(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 September 2002 (06.09.2002)

PCT

(10) International Publication Number
WO 02/068466 A2

(51) International Patent Classification[7]: C07K 14/47, C12N 15/12, 15/63, 5/10, A61K 38/17, C12Q 1/68, G01N 33/68, A61K 48/00

(21) International Application Number: PCT/GB02/00817

(22) International Filing Date: 22 February 2002 (22.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0104440.3 22 February 2001 (22.02.2001) GB

(71) Applicant *(for all designated States except US)*: **OXFORD BIOMEDICA (UK) LIMITED** [GB/GB]; Medawar Centre, Robert Robinson Avenue, The Oxford Science Park, Oxford OX4 4GA (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **WHITE, Jonathan** [GB/GB]; Medawar Centre, Robert Robinson Avenue, The Oxford Science Park, Oxford OX4 4GA (GB). **MUNDY, Christopher, Robert** [AU/GB]; Medawar Centre, Robert Robinson Avenue, The Oxford Science Park, Oxford OX4 4GA (GB). **WARD, Neil, Raymond** [GB/GB]; Medawar Centre, Robert Robinson Avenue, The Oxford Science Park, Oxford OX4 4GA (GB). **KRIGE, David** [GB/GB]; Medawar Centre, Robert Robinson Avenue, The Oxford Science Park, Oxford OX4 4GA (GB). **KINGSMAN, Alan, John** [GB/GB]; Medawar Centre, Robert Robinson Avenue, The Oxford Science Park, Oxford OX4 4GA (GB). **HARRIS, Robert, Alan** [GB/GB]; Medawar Centre, Robert Robinson Avenue, The Oxford Science Park, Oxford OX4 4GA (GB). **BARBER, Robert** [GB/GB]; Medawar Centre, Robert Robinson Avenue, The Oxford Science Park, Oxford OX4 4GA (GB).

(74) Agents: **HALLYBONE, Huw, George** et al.; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/068466 A2

(54) Title: HYPOXIA-REGULATED GENES

(57) Abstract: This invention relates to novel genes and gene products that are implicated in certain disease states. These genes have been identified by a novel method that comprises the steps of comparing: i) the transcriptome or proteome of a first specialised cell type that is implicated in the disease or condition under conditions of both normoxia and hypoxia; with ii) the transcriptome or proteome of a second specialised cell type under conditions of both normoxia and hypoxia; and identifying as a gene implicated in the disease or physiological condition, a gene that is differentially regulated in the two specialised cell types under the two different conditions compared.

HYPOXIA-REGULATED GENES

This invention relates to novel genes and gene products that are implicated in certain disease states.

All publications, patents and patent applications cited herein are incorporated in full by reference.

Over recent years, the so-called "genomics revolution" has allowed access to large portions of whole genomes, including the human genome. The amount of sequence information now available considerably facilitates the analysis of the results of experiments that aim to elucidate the functions of proteins that are expressed in the body. As this information increases in scope and becomes more readily available, the study of the molecular mechanism of disease, and the elucidation of techniques for combating these diseases will be considerably facilitated.

One particular physiological condition that has considerable relevance to human and other animal disease is the cellular response to hypoxia. The term "hypoxia" is intended to refer to an environment of reduced oxygen tension, as compared to the normal physiological environment for a particular organism, which is termed "normoxia".

In a variety of human diseases, cells are exposed to conditions of low oxygen tension, usually as a result of poor oxygen supply to the diseased area. For instance, tissue oxygenation plays a significant regulatory role in both apoptosis and in angiogenesis (Bouck *et al*, 1996, Adv. Cancer Res. 69:135-174; Bunn *et al*, 1996, Physiol. Rev. 76:839-885; Dor *et al*, 1997, Trends Cardiovasc. Med., 7:289-294; Carmeliet *et al*, 1998, Nature 394:485-490). Apoptosis (see Duke *et al*, 1996, Sci. American, 80-87 for review) and growth arrest occur when cell growth and viability are reduced due to oxygen deprivation. Angiogenesis (i.e. blood vessel growth, vascularization), is stimulated when hypooxygenated cells secrete factors that stimulate proliferation and migration of endothelial cells in an attempt to restore oxygen homeostasis (for review see Hanahan *et al*, 1996, Cell, 86:353-364).

Ischaemic disease pathologies involve a decrease in the blood supply to a bodily organ, tissue or body part generally caused by constriction or obstruction of the blood vessels. For example, solid tumours typically have a disorganised blood supply, leading to hypoxic regions. Other disease conditions involving hypoxia include stroke, atherosclerosis, retinopathy, acute renal failure, myocardial infarction, diseases involving infection of the airways (such as cystic fibrosis) and stroke. Therefore, apoptosis and angiogenesis as induced by the ischaemic condition are also considered to be involved in these disease states. It is generally considered that understanding the mechanism by which cells respond to these diseases may be the key to

the disease pathology and thus relevant to disease treatment.

In a different but related approach, it is now recognised that angiogenesis is necessary for tumour growth and that retardation of this process provide a useful tool in controlling malignancy and retinopathies. For example, neoangiogenesis is seen in many forms of retinopathy and in tumour growth. The ability to be able to induce tumourigenic cells to undergo apoptosis is an extremely desirable goal; particularly in the cancer field, it has been observed that apoptosis and angiogenesis-related genes provide potent therapeutic targets. It has also been observed that hypoxia plays a critical role in the selection of mutations that contribute to more severe tumourigenic phenotypes (Graeber *et al.*, 1996 Nature, 379(6560):88-91).

The macrophage is a key cell type of the immune system, which has been shown to play a central role in the pathology of diseases which currently lack suitable treatments. In many of these diseases macrophages are known or suspected to be under the influence of hypoxia. These diseases include solid tumours (including ovarian and breast), dermal wounds, atherosclerosis, rheumatoid arthritis (RA), proliferative retinopathy, cerebral malaria, peripheral arterial disease, chronic occlusive pulmonary disease (Lewis JS *et al.*, 1999 *J.Leukocyte Biol.*. 66: 889-900). In particular, it has been directly demonstrated that hypoxia inducible factor 1α (HIF-1α) becomes activated in macrophages from the joint synovia of rheumatoid arthritis patients, consistent with previous reports of the rheumatoid synovium being hypoxic (Hollander AP *et al.*, 2001 *Arthritis and Rheumatism* 44: 1540-1544). Because the macrophage is accepted as being a key cell type in the pathophysiology and treatment of rheumatoid arthritis (Kinne RW *et al.*, 2000 *Arthritis Research*, 2:189-202), this confirms that the aberrant functions of macrophages in RA are at least partly attributable to its exposure to hypoxia.

Of the genes known in the art to be activated by hypoxia in macrophages, many are involved in inflammation, including cytokines, chemokines, and enzymes involved in prostaglandin E2 synthesis (Lewis JS *et al.*, 1999 *J.Leukocyte Biol.* 66: 889-900). The macrophage is also a key cell type in promoting angiogenesis in response to hypoxia (Knighton DR *et al.*, *Science*. 1983 221:1283-5), and tumour-infiltrated macrophages are key to the promotion of angiogenesis in cancer, and determining prognosis (Leek RD *et al.* 1997, 56:4625-4629). For the above reasons, a complete understanding of the transcriptional response of primary human macrophages, would therefore be of great utility in identifying novel therapeutic strategies for the aforementioned diseases. Strategies at improving the efficiency of this by viral-based overexpression of key transcription factors which mediate the hypoxic response are also

required.

Early in the history of this field it was discovered that a transcription factor, HIF-1alpha, is ubiquitously present in cells and is responsible for the induction of a number of genes in response to hypoxia. This protein is considered a master regulator of oxygen homeostasis (see, for example, Semenza, (1998) Curr. Op. Genetics and Dev. 8:588-594). Although HIF-1α is well known to mediate responses to hypoxia, other transcription factors are also known or suspected to be involved. These include a protein called endothelial PAS domain protein 1 (EPAS1) or HIF-2α, which shares 48% sequence identity with HIF-1α (Tian *et al., Genes Dev.* 1997 Jan 1;11(1):72-82.). Evidence suggests that EPAS1 is especially important in mediating the hypoxia-response in certain cell types, and it is clearly detectable in human macrophages, suggesting a role in this cell type (Griffiths *et al.*, 2000, *Gene Ther.*, 7(3):255-62).

A number of genes regulated by transcription factors such as HIF-1α have been identified that are implicated in the physiological response to hypoxia. However, there remains a great need for the identification of other genes that are implicated in this condition, in order to develop a spectrum of diagnostic and therapeutic agents for use as tools in combating diseases in which hypoxia plays a role. Such genes and the proteins that they encode are candidate targets for antagonist or agonist agents that modulate human disease states. Furthermore, the identified genes are associated with regulatory elements that provide alternative and additional candidate targets for exploitation for the delivery of gene products in a cell-specific fashion. Any genes and regulatory elements identified as having a role in hypoxia may be used directly in therapeutic applications via gene therapy, via recombinant protein methods or via chemical mimetics or as targets for the development of agonists and antagonists such as antibodies, small chemical molecules, peptides, regulatory nucleic acids.

According to the invention, there are provided novel genes and proteins, that have been functionally annotated for the first time. Some of these sequences are only identified as "hypothetical proteins" in the public databases. Each and every one of these sequences forms an embodiment of this aspect of the invention. The invention also includes proteins whose amino acid sequences are encoded by a nucleic acid sequence recited in various ESTs deposited in the public databases or encoded by a gene identified from such an EST. At present, all of these EST sequences are functionally unannotated in the public databases. Each and every one of these sequences forms an embodiment of this aspect of the invention.

One embodiment of the invention thus provides a substantially purified polypeptide, which polypeptide:

i) comprises or consists of the amino acid sequence as recited in any one of SEQ ID Nos: 3, 5, 7, 11, 13, 19, 21, 23 or 27;

ii) comprises or consists of an amino acid sequence encoded by a nucleic acid sequence recited in any one of SEQ ID Nos: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 or 36, or encoded by a gene identified from an EST recited in any one of these SEQ ID Nos;

iii) is a fragment of a polypeptide according to i) or ii), provided that said fragment retains a biological activity possessed by the full length polypeptide of i) or ii), or has an antigenic determinant in common with the polypeptide of i) or ii); or

iv) is a functional equivalent of a polypeptide of i), ii) or (iii).

The polypeptide sequences recited in SEQ ID Nos: 3, 5, 7, 11, 13, 19, 21, 23 and 27 were, prior to the present disclosure, unannotated in the literature and public sequence databases. Accordingly, until now, no biological function has been attributed to these polypeptide sequences; each of these sequences is generally labelled in the databases as a "hypothetical protein". A novel method termed "Smartomics", described herein and in co-pending International patent application PCT/GB01/00758, has now elucidated a biological function for these polypeptides, in that they have been found to be differentially regulated under physiological conditions of hypoxia.

For the purposes of this document, the term "hypoxia" should be taken to mean an environment of oxygen tension such that the oxygen content is between about 5% and 0.1% (v/v). In most cases, hypoxic tissue will have an oxygen content that is less than or equal to about 2%. The term "normoxia" should be taken to mean conditions comprising a normal level of oxygen for the environment concerned. Normoxic tissue typically has an oxygen content above about 5%.

These discoveries allow the development of regulators, such as small drug molecules, that affect the activity of these polypeptides, so allowing diseases and physiological conditions that are caused by hypoxia, or in which hypoxia has been implicated, to be treated. These discoveries also allow the development of diagnostic agents that are suitable for the detection of hypoxia in biological tissues and, through the identification of mutations and polymorphisms (such as SNPs) within genes coding for the proteins implicated herein, allows the assessment of an individual's risk of being susceptible to diseases and physiological conditions in which hypoxia is implicated.

The biological activity of polypeptides whose sequences are listed in SEQ ID Nos: 3, 5, 7, 11, 13, 19, 21, 23 and 27 has been found to be hypoxia-regulated. The expression of these polypeptides has been found to be induced under conditions of hypoxia. By "hypoxia-induced" is meant that the polypeptide is expressed at a higher level when a cell is exposed to hypoxia conditions as compared to its expression level under conditions of normoxia. The term "hypoxia-repressed" as used herein is intended to mean that the polypeptide is expressed at a lower level when a cell is exposed to hypoxia conditions as compared to its expression level under conditions of normoxia.

In addition, polypeptides whose amino acid sequence is encoded by a nucleic acid sequence recited in any one of SEQ ID Nos.: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 and 36 or encoded by a gene identified from an EST recited in any one of these SEQ ID Nos have been found to be hypoxia-regulated.

The polypeptide sequences whose amino acid sequence is encoded by a nucleic acid sequence recited in SEQ ID Nos: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 and 36, or whose amino acid sequence is encoded by a gene identified from an EST recited in any one of these SEQ ID Nos., were also, prior to the present disclosure, unannotated in the literature and public sequence databases, meaning that until now, no biological function has been attributed to these polypeptide sequences.

The sequences in this group fall into a number of different categories. The first of these are cDNA clones, for which a protein sequence has not been predicted by the depositor. A second category is expressed sequence tag (EST) sequences that are represented in the UniGene database (http://www.ncbi.nlm.nih.gov/UniGene/), which contain modest or weak homology to known proteins when translated. ESTs are single-pass sequence files of the 5' region of an organism's expressed genome as accessed via a force-cloned cDNA library. EST sequences tend to be short and as a general rule are error-prone. UniGene (see http://www.ncbi.nlm.nih.gov/Web/Newsltr/aug96.html for review) is an experimental system for automatically partitioning these EST sequences into a non-redundant set of gene-oriented clusters. Each UniGene cluster contains sequences that are considered to represent a unique gene, as well as related information such as the tissue types in which the gene has been expressed and map location. A third category of hits identified by the methods described herein involves EST sequences that are contained in Unigene clusters, but which are not annotated and exhibit no homologies to proteins contained in the public databases. The fourth and final category encompasses singleton EST sequence entries that are not incorporated as entries in the Unigene database and that only appear as single entries in the public databases.

A biological function has now been attributed to the polypeptides that are encoded by genes incorporating cDNA and EST sequences that fall into the four categories set out above, in that these sequences have been found to be differentially regulated under physiological conditions of hypoxia. Such polypeptides may have an amino acid sequence that is encoded by a nucleic acid sequence recited in any one of SEQ ID Nos: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 and 36. However, the EST sequences in particular may not be part of the actual coding sequence for a gene, often representing regulatory regions of the gene, or regions that are transcribed, but not translated into polypeptide. Accordingly, this aspect of the invention also includes polypeptides that are encoded by a gene identified from an EST recited in any one of SEQ ID Nos: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 and 36.

Polypeptides of this aspect of the invention are intended to include fragments of polypeptides according to i) or ii) as defined above, provided that the fragment retains a biological activity that is possessed by the full length polypeptide of i) or ii), or has an antigenic determinant in common with the polypeptide of i) or ii). As used herein, the term "fragment" refers to a polypeptide having an amino acid sequence that is the same as part, but not all, of an amino acid sequence as recited in any one of SEQ ID Nos: 3, 5, 7, 11, 13, 19, 21, 23 and 27, an amino acid sequence that is encoded by a nucleic acid sequence recited in any one of SEQ ID Nos. 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 and 36, or an amino acid sequence that is encoded by a gene that is linked to a nucleic acid sequence recited in any one of these SEQ ID Nos. The fragments should comprise at least *n* consecutive amino acids from the sequence and, depending on the particular sequence, *n* preferably is 7 or more (for example, 8, 10, 12, 14, 16, 18, 20 or more). Small fragments may form an antigenic determinant.

Such fragments may be isolated fragments, that are not part of or fused to other amino acids or polypeptides, or they may be comprised within a larger polypeptide, of which they form a part or region. When comprised within a larger polypeptide, a fragment of the invention most preferably forms a single continuous region. For instance, certain preferred embodiments relate to a fragment having a pre – and/or pro- polypeptide region fused to the amino terminus of the fragment and/or an additional region fused to the carboxyl terminus of the fragment. However, several fragments may be comprised within a single larger polypeptide.

The polypeptides of the present invention or their immunogenic fragments (comprising at least one antigenic determinant) can be used to generate ligands, such as polyclonal or monoclonal antibodies, that are immunospecific for the polypeptides. Such antibodies may be employed to isolate or to identify clones that express a polypeptide according to the invention or, for example, to purify the polypeptide by affinity chromatography. Such antibodies may also be

employed as diagnostic or therapeutic aids, amongst other applications, as will be apparent to the skilled reader.

The term "immunospecific" means that an antibody has substantially greater affinity for a polypeptide according to the invention than their affinity for related polypeptides. As used herein, the term "antibody" is intended to include intact molecules as well as fragments thereof, such as Fab, $F(ab')_2$ and scFv, which are capable of binding to the antigenic determinant in question.

The invention also includes functional equivalents of a polypeptide of i), ii) or (iii) as recited above. A functionally-equivalent polypeptide according to this aspect of the invention may be a polypeptide that is homologous to a polypeptide whose sequence is explicitly recited herein. Two polypeptides are said to be "homologous" if the sequence of one of the polypeptides has a high enough degree of identity or similarity to the sequence of the other polypeptide. "Identity" indicates that at any particular position in the aligned sequences, the amino acid residue is identical between the sequences. "Similarity" indicates that, at any particular position in the aligned sequences, the amino acid residue is of a similar type between the sequences. Degrees of identity and similarity can be readily calculated according to methods known in the art (see, for example, Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing. Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993).

Typically, greater than 50% identity between two polypeptides is considered to be an indication of functional equivalence, provided that either the biological activity of the polypeptide is retained or the polypeptides possess an antigenic determinant in common. Preferably, a functionally equivalent polypeptide according to this aspect of the invention exhibits a degree of sequence identity with a polypeptide sequence explicitly identified herein, or with a fragment thereof, of greater than 50%. More preferred polypeptides have degrees of identity of greater than 60%, 70%, 80%, 90%, 95%, 98% or 99%, respectively.

Functionally-equivalent polypeptides according to the invention are therefore intended to include natural biological variants (for example, allelic variants or geographical variations within the species from which the polypeptides are derived) and mutants (such as mutants containing amino acid substitutions, insertions or deletions) of the polypeptides whose sequences are explicitly recited herein. Such mutants may include polypeptides in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or

may not be one encoded by the genetic code. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; among the basic residues Lys and Arg; or among the aromatic residues Phe and Tyr.

Particularly preferred are variants in which several, i.e. between 5 and 10, 1 and 5, 1 and 3, 1 and 2 or just 1 amino acids are substituted, deleted or added in any combination. Especially preferred are silent substitutions, additions and deletions, which do not alter the properties and activities of the protein. Also especially preferred in this regard are conservative substitutions. "Mutant" polypeptides also include polypeptides in which one or more of the amino acid residues includes a substituent group.

According to a further aspect of the invention, there is provided a purified and isolated nucleic acid molecule that encodes a polypeptide according to any one of the aspects of the invention discussed above. Such a nucleic acid molecule may consist of a nucleic acid sequence as recited in any one of SEQ ID Nos. 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 and 36, or form a redundant equivalent or fragment thereof. This aspect of the invention also includes a purified nucleic acid molecule which hydridizes under high stringency conditions with a nucleic acid molecule as described above.

According to a further aspect of the invention, there is provided an expression vector that contains a purified and isolated nucleic acid molecule according to the aspects of the invention described above. The invention also incorporates a delivery vehicle, such as a liposome, comprising a nucleic acid according to the above-described aspects of the invention.

In a further aspect, the invention provides a host cell transformed with a vector of the above-described aspect of the invention.

In a still further aspect, the invention provides a ligand that binds specifically to a polypeptide according to the above-described aspects of the invention. The ligand may be an antagonist ligand that inhibits the biological activity of the polypeptide, or may be an agonist ligand that activates the hypoxia-induced activity of the polypeptide to augment or potentiate a hypoxia-induced activity. The term "ligand", as used herein, is used broadly and is intended to encompass natural or modified substrates, enzymes, receptors, small organic molecules such as small natural or synthetic organic molecules of up to 2000Da, preferably 800Da or less, peptidomimetics, inorganic molecules, peptides, polypeptides, antibodies, structural or functional mimetics of the aforementioned and all other types of chemical entities that act in the manner specified.

In a still further aspect of the invention, there is provided a ligand which binds specifically to, and which preferably inhibits the hypoxia-induced activity of, a polypeptide according to any one of the above-described aspects of the invention. Such a ligand may, for example, be an antibody that is immunospecific for the polypeptide in question.

According to a further aspect, the invention provides a polypeptide, a nucleic acid molecule, vector or ligand as described above, for use in therapy or diagnosis of a disease or abnormal physiological condition. Preferably, the disease or abnormal physiological condition that is affected by hypoxia; examples of such diseases include cancer, ischaemic conditions (such as stroke, coronary arterial disease, peripheral arterial disease), reperfusion injury, retinopathy, neonatal stress, preeclampsia, atherosclerosis, inflammatory conditions (including rheumatoid arthritis), diseases involving infections of the airways (such as cystic fibrosis) and wound healing. The undesired cellular process involved in said diseases might include, but is not restricted to; tumorigenesis, angiogenesis, apoptosis, inflammation or erythropoiesis. The undesired biochemical processes involved in said cellular processes might include, but is not restricted to, glycolysis, gluconeogenesis, glucose transportation, catecholamine synthesis, iron transport or nitric oxide synthesis.

According to the invention, a protein known to be implicated in the biological response to hypoxia has now been found to be regulated by HIF1α. Certain functions of this protein are known, meaning that these functions have been annotated in the public databases. The sequence of this protein is presented in SEQ ID No: 37; the encoding gene sequence is presented as SEQ ID No: 38.

According to a further aspect of the invention, there is provided a substantially purified polypeptide, which polypeptide:

i) comprises the amino acid sequence as recited in any one of SEQ ID Nos: 3, 5, 7, 11, 13, 19, 21, 23, 27 or 37;

ii) has an amino acid sequence encoded by a nucleic acid sequence recited in any one of SEQ ID Nos: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36 or 38, or encoded by a gene identified from an EST recited in any one of these SEQ ID Nos;

iii) is a fragment of a polypeptide according to i) or ii), provided that said fragment retains a biological activity possessed by the full length polypeptide of i) or ii), or has an antigenic determinant in common with the polypeptide of i) or ii); or

iv) is a functional equivalent of a polypeptide of i), ii) or (iii);

for use in the diagnosis or therapy of tumourigenesis, angiogenesis, apoptosis, the biological response to hypoxia conditions, or a hypoxic-associated pathology.

The invention also provides a purified and isolated nucleic acid molecule that encodes a polypeptide according to this aspect of the invention, for use in the diagnosis or therapy of tumourigenesis, angiogenesis, apoptosis, the biological response to hypoxia conditions, or a hypoxic-associated pathology. The sequences of these molecules are provided in SEQ ID Nos.: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36 and 38. As described above for the EST nucleic acid sequences annotated herein, this aspect of the invention includes redundant equivalents and fragments of the sequences explicitly recited in SEQ ID Nos.: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36 and 38, and purified nucleic acid molecules which hybridize under high stringency conditions with such nucleic acid molecules, and vectors containing such nucleic acid molecules for use in the diagnosis or therapy of tumourigenesis, angiogenesis, apoptosis, the biological response to hypoxia conditions, or a hypoxic-associated pathology.

This aspect of the invention also includes ligands which bind specifically to, and which preferably inhibit the hypoxia-induced activity of, a polypeptide listed in SEQ ID Nos.: 3, 5, 7, 11, 13, 19, 21, 23, 27 or 37, or encoded by a nucleic acid sequence recited in any one of SEQ ID Nos: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36 or 38, or encoded by a gene identified from an EST recited in any one of these SEQ ID Nos, for use in the diagnosis or therapy of tumourigenesis, angiogenesis, apoptosis, the biological response to hypoxia conditions, or a hypoxic-associated pathology.

The invention also provides a pharmaceutical composition suitable for modulating hypoxia and/or ischaemia, comprising a therapeutically-effective amount of a a polypeptide, a nucleic acid molecule, vector or ligand as described above, in conjunction with a pharmaceutically-acceptable carrier.

The invention also provides a vaccine composition comprising a polypeptide, or a nucleic acid molecule as described above.

The invention also provides a method of treating a disease in a patient in need of such treatment by administering to a patient a therapeutically effective amount of a polypeptide, a nucleic acid molecule, vector, ligand or pharmaceutical composition as described above. For diseases in which the expression of the natural gene or the activity of the polypeptide is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand, compound or composition administered to the

patient should be an agonist. For diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an antagonist. By the term "agonist" is meant herein, any polypeptide, peptide, synthetic molecule or organic molecule that functions as an activator, by increasing the effective biological activity of a polypeptide, for example, by increasing gene expression or enzymatic activity. By the term "antagonist" is meant herein, any polypeptide, peptide, synthetic molecule or organic molecule that functions as an inhibitor, by decreasing the effective biological activity of the gene product, for example, by inhibiting gene expression of an enzyme or a pharmacological receptor.

The invention also provides a polypeptide, nucleic acid molecule, vector, ligand or pharmaceutical composition according to any one of the above-described aspects of the invention, for use in the manufacture of a medicament for the treatment of a hypoxia-regulated condition.

The invention also provides a method of monitoring the therapeutic treatment of disease or physiological condition in a patient, comprising monitoring over a period of time the level of expression or activity of a polypeptide, nucleic acid molecule, vector or ligand according to any one of the above-described aspects of the invention in tissue from said patient, wherein altering said level of expression or activity over the period of time towards a control level is indicative of regression of said disease or physiological condition.

The invention also provides a method of providing a hypoxia regulating gene, an apoptotic or an angiogenesis regulating gene by administering directly to a patient in need of such therapy an expressible vector comprising expression control sequences operably linked to one or more of the nucleic acid molecules as described above.

The invention also provides a method of diagnosing a hypoxia-regulated condition in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide according to any one of the aspects of the invention described above in tissue from said patient and comparing said level of expression or activity to a control level, wherein a level that is different to said control level is indicative of the hypoxia-related condition.

Such a method of diagnosis may be carried out *in vitro*. One example of a suitable method comprises the steps of: (a) contacting a ligand as described above with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

A further example of a suitable method may comprises the steps of: a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule whose sequence is recited in any one of SEQ ID Nos.: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 and 36, and the probe; b) contacting a control sample with said probe under the same conditions used in step a); and c) detecting the presence of hybrid complexes in said samples; wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid complex in the control sample is indicative of the hypoxia-related condition.

A still further example of a suitable method may comprise the steps of: a) contacting a sample of nucleic acid from tissue of the patient with a nucleic acid primer under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule whose sequence is recited in any one of SEQ ID Nos.: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 and 36, and the primer; b) contacting a control sample with said primer under the same conditions used in step a); c) amplifying the sampled nucleic acid; and d) detecting the level of amplified nucleic acid from both patient and control samples; wherein detection of levels of the amplified nucleic acid in the patient sample that differ significantly from levels of the amplified nucleic acid in the control sample is indicative of the hypoxia-related condition.

A still further example of a suitable method may comprised the steps of: a) obtaining a tissue sample from a patient being tested for the hypoxia-related condition; b) isolating a nucleic acid molecule according to any one of the above-described aspects of the invention from said tissue sample; and c) diagnosing the patient for the hypoxia-related condition by detecting the presence of a mutation which is associated with the hypoxia-related condition in the nucleic acid molecule as an indication of the hypoxia-related condition. This method may comprise the additional step of amplifying the nucleic acid molecule to form an amplified product and detecting the presence or absence of a mutation in the amplified product.

Particular hypoxia-related conditions that may be diagnosed in this fashion include cancer, ischaemia, reperfusion, retinopathy, neonatal stress, preeclapmsia, atherosclerosis, rheumatoid arthritis, cardiac arrest or stroke, for example, caused by a disorder of the cerebral, coronary or peripheral circulation.

In a further aspect, the invention provides a method for the identification of a compound that is effective in the treatment and/or diagnosis of a hypoxia-regulated condition, comprising contacting a polypeptide, nucleic acid molecule, or ligand according to any one of the above-described aspects of the invention with one or more compounds suspected of possessing

binding affinity for said polypeptide, nucleic acid molecule or ligand, and selecting a compound that binds specifically to said nucleic acid molecule, polypeptide or ligand.

According to a still further aspect of the invention, there is provided a kit useful for diagnosing a hypoxia-regulated condition, comprising a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to any one of the aspects of the invention described above; a second container containing primers useful for amplifying said nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of the hypoxia-regulated condition. The kit may additionally comprise a third container holding an agent for digesting unhybridised RNA.

To facilitate in the diagnosis of the hypoxia-regulated condition using one of the methods outlined above, in a further aspect, the invention provides an array of at least two nucleic acid molecules, wherein each of said nucleic acid molecules either corresponds to the sequence of, is complementary to the sequence of, or hybridises specifically to a nucleic acid molecule according to any one of the aspects of the invention described above. Such an array may contain nucleic acid molecules that either correspond to the sequence of, are complementary to the sequence of, or hybridise specifically to at least 1-18 or more of the nucleic acid molecules implicated in a hypoxia-regulated condition as recited above. The nucleic acid molecules on the array may consist of oligonucleotides of between twelve and fifty nucleotides, more preferably, between forty and fifty nucleotides. Alternatively, the nucleic acid molecules on the array may consist of PCR-amplified cDNA inserts where the nucleic acid molecule is between 300-2000 nucleotides.

In a related aspect, again useful for diagnosis, the invention provides an array of antibodies, comprising at least two different antibody species, wherein each antibody species is immunospecific with a polypeptide implicated in a hypoxia-regulated condition as described above. The invention also provides an array of polypeptides, comprising at least two polypeptide species as recited above, wherein each polypeptide species is implicated in a hypoxia-regulated condition, or is a functional equivalent variant or fragment thereof.

Kits useful in the diagnostic methods of the invention may comprise such nucleic acid, antibody and/or polypeptide arrays.

According to the invention, a kit may also comprise one or more antibodies that bind to a polypeptide as recited above, and a reagent useful for the detection of a binding reaction between said antibody and said polypeptide.

According to a still further aspect of the invention, there is provided a genetically-modified non-human animal that has been transformed to express higher, lower or absent levels of a polypeptide according to any one of the aspects of the invention described above. Preferably, said genetically-modified animal is a transgenic or knockout animal.

The invention also provides a method for screening for a compound effective to treat a hypoxia-regulated condition, by contacting a non-human genetically-modified animal as described above with a candidate compound and determining the effect of the compound on the physiological state of the animal.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA technology and immunology, which are within the skill of those working in the art.

Most general molecular biology, microbiology recombinant DNA technology and immunological techniques can be found in Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (1989) Cold Harbor-Laboratory Press, Cold Spring Harbor, N.Y. or Ausubel *et al.*, Current protocols in molecular biology (1990) John Wiley and Sons, N.Y.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs.

A. Polypeptides .

Polypeptides

The term "polypeptide" as used herein, refers to a chain (may be branched or unbranched) of two or more amino acids linked to each other by means of a peptide bond or modified peptide bond (isosteres). The term polypeptide encompasses but is not limited to oligopeptides, peptides and proteins. The polypeptide of the invention may additionally be either in a mature protein form or in a pre-, pro- or prepro-protein form that requires subsequent cleavage for formation of the active mature protein. The pre-, pro-, prepro- part of the protein is often a leader or secretory sequence but may also be an additional sequence added to aid protein purification (for example, a His tag) or to conform a higher stability to the protein.

A polypeptide according to the invention may also include modified amino acids, that is, amino acids other than those 20 that are gene-encoded. This modification may be a result of natural processes such as post-translational processing or by chemical modification. Examples of modifications include acetylation, acylation, amidation, ADP-ribosylation, arginylation,

attachment of a lipid derivative or phosphatidylinositol, γ-carboxylation, covalent attachment of a flavin or haeme moiety, a nucleotide or nucleotide derivative, cyclisation, demethylation, disulphide bond formation, formation of covalent cross-links, formylation, glycosylation, GPI anchor formation, hydroxylation, iodination, lipid attachment, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemisation, selenoylation, sulphation, and ubiquitination. Modification of the polypeptide can occur anywhere within the molecule including the backbone, the amino acid side-chains or at the N- or C-terminals.

A polypeptide according to the invention may either be isolated from natural sources (for example, purified from cell culture), or be a recombinantly produced polypeptide, or a synthetically produced polypeptide or a combination of all the above.

Antibodies

A polypeptide according to the invention, its functional equivalents and/or any immunogenic fragments derived from the polypeptide may be used to generate ligands including immunospecific monoclonal or polyclonal antibodies, or antibody fragments. These antibodies can then be used to isolate or identify clones expressing the polypeptide of the invention or to purify the polypeptide by affinity chromatography. Further uses of these immunospecific antibodies may include, but are not limited to, diagnostic, therapeutic or general assay applications. Examples of assay techniques that employ antibodies are immunoassays, radioimmunoassays (RIA) or enzyme linked immunosorbent assay (ELISA). In these cases, the antibodies may be labelled with an analytically-detectable reagent including radioisotopes, a fluorescent molecule or any reporter molecule.

The term "immunospecific" as used herein refers to antibodies that have a substantially higher affinity for a polypeptide of this invention compared with other polypeptides. The term "antibody" as used herein refers to a molecule that is produced by animals in response to an antigen and has the particular property of interacting specifically with the antigenic determinant that induced its formation. Fragments of the aforementioned molecule such as Fab, $F(ab')_2$ and scFv, which are capable of binding the antigen determinant, are also included in the term "antibody". Antibodies may also be modified to make chimeric antibodies, where non-human variable regions are joined or fused to human constant regions (for example, Liu *et al.*, PNAS, USA, 84, 3439 (1987)). Particularly, antibodies may be modified to make them less immunogenic to an individual in a process such as humanisation (see, for example, Jones *et al.*, Nature, 321, 522 (1986); Verhoeyen *et al.*, Science, 239, 1534 (1988); Kabat *et al.*, J. Immunol., 147, 1709 (1991); Queen *et al.*, PNAS, USA, 86, 10029 (1989); Gorman *et al.*,

PNAS, USA, 88, 34181 (1991) and Hodgson *et al.*, Bio/Technology, 9, 421 (1991)). The term "humanised antibody", as used herein, refers to antibody molecules in which the amino acids of the CDR (complementarity-determining region) and selected other regions in the variable domains of the heavy and/or light chains of a non-human donor antibody have been substituted with the equivalent amino acids of a human antibody. The humanised antibody therefore closely resembles a human antibody, but has the binding ability of the donor antibody. Antibodies may also have a "bispecific" nature, that is, the antibody has two different antigen binding domains, each domain being directed against a different epitope.

Specific polyclonal antibodies may be made by immuno-challenging an animal with a polypeptide of this invention. Common animals used for the production of antibodies include the mouse, rat, chicken, rabbit, goat and horse. The polypeptide used to immuno-challenge the animal may be derived by recombinant DNA technology or may be chemically-synthesised. In addition, the polypeptide may be conjugated to a carrier protein. Commonly used carriers to which the polypeptides may be conjugated include, but are not limited to BSA (bovine serum albumin), thyroglobulin and keyhole limpet haemocyanin. Serum from the immuno-challenged animal is collected and treated according to known procedures, for example, by immunoaffinity chromatography.

Specific monoclonal antibodies can generally be made by methods known to one skilled in the art (see for example, Kohler, G. and Milstein, C., Nature 256, 495-497 (1975); Kozbor *et al.*, Immunology Today 4: 72 (1983); Cole *et al.*, 77-96 in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. (1985) and Roitt, I. *et al.*, Immunology, 25.10, Mosby-Year Book Europe Limited (1993)). Panels of monoclonal antibodies produced against the polypeptides of the invention can be screened for various properties, i.e., for isotype, epitope, affinity, etc. against which they are directed. Alternatively, genes encoding the monoclonal antibodies of interest may be isolated from hybridomas, for instance using PCR techniques known in the art, and cloned and expressed in appropriate vectors.

Phage display technology may be utilised to select the genes encoding the antibodies that have exhibited an immunospecific response to the polypeptides of the invention (see McCafferty, J., *et al.*, (1990), Nature 348, 552-554; Marks, J. *et al.*, (1992) Biotechnology 10, 779-783).

Ligands

The polypeptides of the invention may also be used to search for interacting ligands. Methods for doing this include the screening of a library of compounds (see Coligan *et al.*, Current

Protocols in Immunology 1(2); Chapter 5 (1991), isolating the ligands from cells, isolating the ligands from a cell-free preparation or natural product mixtures. Ligands to the polypeptide may activate (agonise) or inhibit (antagonise) its activity. Alternatively, compounds may affect the levels of the polypeptide present in the cell, including affecting gene expression and/or mRNA stability.

Ligands to the polypeptide form a further aspect of the invention, as discussed in more detail above. Preferred "antagonist" ligands include those that bind to the polypeptide of this invention and strongly inhibit any activity of the polypeptide. Preferred "agonist" ligands include those that bind to the polypeptide and strongly induce activity of the polypeptide of this invention or increases substantially the level of the polypeptide in the cell. As defined above, the term "agonist" is meant to include any polypeptide, peptide, synthetic molecule or organic molecule that functions as an activator, by increasing the effective biological activity of a polypeptide, for example, by increasing gene expression or enzymatic activity. The term "antagonist" is meant to include any polypeptide, peptide, synthetic molecule or organic molecule that functions as an inhibitor, by decreasing the effective biological activity of the gene product, for example, by inhibiting gene expression of an enzyme or a pharmacological receptor.

Ligands to a polypeptide according to the invention may come in various forms, including natural or modified substrates, enzymes, receptors, small organic molecules such as small natural or synthetic organic molecules of up to 2000Da, preferably 800Da or less, peptidomimetics, inorganic molecules, peptides, polypeptides, antibodies, structural or functional mimetics of the aforementioned.

B. Nucleic acid molecules

Preferred nucleic acid molecules of the invention are those which encode the polypeptide sequences recited in any one of SEQ ID Nos. 3, 5, 7, 11, 13, 19, 21, 23 or 27, or which encode polypeptides encoded by a nucleic acid sequence recited in any one of SEQ ID Nos: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 or 36, or encoded by a gene identified from an EST recited in any one of these SEQ ID Nos. Examples of such nucleic acid molecules include those listed in SEQ ID Nos. 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 and 36, homologous nucleic acids and nucleic acids that are complementary to these nucleic acid molecules. Nucleic acid molecules of this aspect of the invention may be used in numerous methods and applications, as described generally herein. A nucleic acid molecule preferably comprises of at least *n* consecutive nucleotides from any one of the sequences disclosed in

SEQ ID Nos.: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 and 36, where *n* is 10 or more. A nucleic acid molecule of the invention also includes sequences that are complementary to the nucleic acid molecule described above (for example, for antisense or probing purposes).

A nucleic acid molecule according to this aspect of the invention may be in the form of RNA, such as mRNA, DNA, such as cDNA, synthetic DNA or genomic DNA. The nucleic acid molecule may be double-stranded or single-stranded. The single-stranded form may be the coding (sense) strand or the non-coding (antisense) strand. A nucleic acid molecule may also comprise an analogue of DNA or RNA, including, but not limited to modifications made to the backbone of the molecule, such as, for example, a peptide nucleic acid (PNA). The term "PNA" as used herein, refers to an antisense molecule that comprises an oligonucleotide of at least five nucleotides in length linked to a peptide backbone of amino acid residues, preferably ending in lysine. The terminal lysine confers solubility to the composition. PNAs may be pegylated to extend their lifespan in a cell, where they preferentially bind complementary single-stranded DNA and RNA and stop transcript elongation (Nielsen, P.E. *et al.* (1993) Anticancer Drug Des. 8:53-63).

A nucleic acid molecule according to this aspect of the invention can be isolated by cloning, purification or separation of the molecule directly from a particular organism, or from a library, such as a genomic or cDNA library. The molecule may also be synthesised, for example, using chemical synthetic techniques such as solid phase phosphoramidite chemical synthesis. RNA may be synthesized *in vitro* or *in vivo* by transcription of the relevant DNA molecule.

Due to the degeneracy of the genetic code, differing nucleic acid sequences may encode the same polypeptide (or mature polypeptide). Thus, nucleic acid molecules included in this aspect of the invention include any molecule comprising a variant of the sequence explicitly recited. Such variants may include variant nucleic acid molecules that code for the same polypeptide (or mature polypeptide) as that explicitly identified, that code for a fragment of the polypeptide, that code for a functional equivalent of the polypeptide or that code for a fragment of the functional equivalent of the polypeptide. Also included in this aspect of the invention, are variant nucleic acid molecules that are derived from nucleotide substitutions, deletions, rearrangements or insertions or multiple combinations of the aforementioned. Such molecules may be naturally occurring variants, such as allelic variants, non-naturally occurring variants such as those created by chemical mutagenesis, or variants isolated from a species, cell or organism type other than the type from which the sequence explicitly identified originated.

Variant nucleic acid molecules may differ from the nucleic acid molecule explicitly recited in a coding region, non-coding region or both these regions.

Nucleic acid molecules may also include additional nucleic acid sequence to that explicitly recited, for example, at the 5' or 3' end of the molecule. Such additional nucleic acids may encode for a polypeptide with added functionality compared with the original polypeptide whose sequence is explicitly identified herein. An example of this would be an addition of a sequence that is heterologous to the original nucleic acid sequence, to encode a fusion protein. Such a fusion protein may be of use in aiding purification procedures or enabling techniques to be carried out where fusion proteins are required (such as in the yeast two hybrid system). Additional sequences may also include leader or secretory sequences such as those coding for pro-, pre- or prepro- polypeptide sequences. These additional sequences may also include non-coding sequences that are transcribed but not translated including ribosome binding sites and termination signals.

A nucleic acid molecule of the invention may include molecules that are at least 70% identical over their entire length to a nucleic acid molecule as explicitly identified herein in SEQ ID Nos.: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 or 36. Preferably, a nucleic acid molecule according to this aspect of the invention comprises a region that is at least 80% identical over its entire length to a nucleic acid molecule as explicitly identified herein in these SEQ ID Nos., preferably at least 90%, more preferably at least 95% and most preferably at least 98% or 99% identical. Further preferred embodiments include nucleic acid molecules that encode polypeptides that retain substantially the same biological function or activity as the polypeptide explicitly identified herein.

The nucleic acid molecules of the invention can also be engineered using methods generally known in the art. These methods include but are not limited to DNA shuffling; random or non-random fragmentation (by restriction enzymes or shearing methods) and reassembly of fragments; insertions, deletions, substitutions and rearrangements of sequences by site-directed mutagenesis (for example, by PCR). These alterations may be for a number of reasons including for ease of cloning (such as introduction of new restriction sites), altering of glycosylation patterns, changing of codon preferences, splice variants changing the processing, and/or expression of the gene product (the polypeptide) in general or creating fusion proteins (see above).

Hybridisation

Nucleic acid molecules of the invention may also include antisense molecules that are partially complementary to a nucleic acid molecule as explicitly identified herein in SEQ ID Nos.: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 or 36, and which therefore will hybridise to the encoding nucleic acid molecules. These antisense molecules, including oligonucleotides, can be designed to recognise, specifically bind to and prevent transcription of a target nucleic acid encoding a polypeptide of the invention, as will be known by those of ordinary skill in the art (see Cohen, J.S., Trends in Pharm. Sci., 10, 435 (1989), Okano, J. Neurochem. 56, 560 (1991); O'Connor, J. Neurochem 56, 560 (1991); Lee *et al.*, Nucleic Acids Res 6, 3073 (1979); Cooney *et al.*, Science 241, 456 (1988); Dervan *et al.*, Science 251, 1360 (1991).

The term "hybridisation" used herein refers to any process by which a strand of nucleic acid binds with a complementary strand of nucleic acid by hydrogen bonding, typically forming Watson-Crick base pairs. As carried out *in vitro*, one of the nucleic acid populations is usually immobilised to a surface, whilst the other population is free. The two molecule types are then placed together under conditions conducive to binding.

The phrase "stringency of hybridisation" refers to the percentage of complementarity that is needed for duplex formation. "Stringency" thus refers to the conditions in a hybridization reaction that favour the association of very similar molecules over association of molecules that differ. Conditions can therefore exist that allow not only nucleic acid strands with 99-100% complementarity to hybridise, but sequences with lower complementarity (for example, 50%) to also hybridise. High stringency hybridisation conditions are defined herein as overnight incubation at 42°C in a solution comprising 50% formamide, 5XSSC (150mM NaCl, 15mM trisodium citrate), 50mM sodium phosphate (pH7.6), 5x Denhardts solution, 10% dextran sulphate, and 20 microgram/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1X SSC at approximately 65°C. Low stringency conditions involve the hybridisation reaction being carried out at 35°C (see Sambrook *et al.* [supra]). Preferably, the conditions used for hybridization are those of high stringency.

Some *trans*- and *cis*-acting factors that may affect the binding of two complementary strands include strand length, base composition (GC pairs have an extra hydrogen bond and are thus require more energy to separate than AT pairs) and the chemical environment. The presence of monovalent cations (such as Na^+) stabilises duplex formation whereas chemical denaturants such as formamide and urea destabilise the duplex by disruption of the hydrogen bonds. Use of compounds such as polyethylene glycol (PEG) can increase reassociation speeds by increasing

overall DNA concentration in aqueous solution by abstracting water molecules. Denhardt's reagent or BLOTTO are chemical agents often added to block non-specific attachment of the liquid phase to the solid support. Increasing the temperature will also increase the stringency of hybridisation, as will increasing the stringency of the washing conditions following hybridisation (Sambrook *et al.* [*supra*]).

Numerous techniques exist for effecting hybridisation of nucleic acid molecules. Such techniques usually involve one of the nucleic acid populations being labelled. Labelling methods include, but are not limited to radiolabelling, fluorescence labelling, chemiluminescent or chromogenic labelling or chemically coupling a modified reporter molecule to a nucleotide precursor such as the biotin-streptavidin system. This can be done by oligolabelling, nick-translation, end-labelling or PCR amplification using a labelled polynucleotide. Labelling of RNA molecules can be achieved by cloning the sequences encoding the polypeptide of the invention into a vector specifically for this purpose. Such vectors are known in the art and may be used to synthesise RNA probes *in vitro* by the addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labelled nucleotides.

Various kits are commercially available that allow the labelling of molecules. Examples include those made by Pharmacia & Upjohn (Kalamazoo, MI); Promega (Madison WI); and the U.S. Biochemical Corp. (Cleveland, OH). Hybridisation assays include, but are not limited to dot-blots, Southern blotting, Northern blotting, chromosome *in situ* hybridisation (for example, FISH [fluorescence *in situ* hybridisation]), tissue *in situ* hybridisation, colony blots, plaque lifts, gridded clone hybridisation assays, DNA microarrays and oligonucleotide microarrays. These hybridisation methods and others, may be used by a skilled artisan to isolate copies of genomic DNA, cDNA, or RNA encoding homologous or orthologous proteins from other species.

The invention therefore also embodies a process for detecting a nucleic acid molecule according to the invention, comprising the steps of: (a) contacting a nucleic probe with a biological sample under hybridising conditions to form duplexes: and (b) detecting any such duplexes that are formed. The term "probe" as used herein refers to a nucleic acid molecule in a hybridisation reaction whose molecular identity is known and is designed specifically to identify nucleic acids encoding homologous genes in other species. Usually, the probe population is the labelled population, but this is not always the case, as for example, in a reverse hybridisation assay.

One example of a use of a probe is to find nucleic acid molecules with an equivalent function to those that are explicitly identified herein, or to identify additional family members in the same or other species. This can be done by probing libraries, such as genomic or cDNA libraries, derived from a source of interest, such as a human, a non-human animal, other eukaryote species, a plant, a prokaryotic species or a virus. The probe may be natural or artificially designed using methods recognised in the art (for example, Ausubel *et al.*, [*supra*]). A nucleic acid probe will preferably possess greater than 15, more preferably greater than 30 and most preferably greater than 50 contiguous bases complementary to a nucleic acid molecule explicitly identified herein.

In many cases, isolated DNA from cDNA libraries will be incomplete in the region encoding the polypeptide, normally at the 5' end. Methods available for subsequently obtaining full-length cDNA sequence include RACE (rapid amplification of cDNA ends) as described by Frohman *et al.*, (Proc. Natl. Acad. Sci. USA 85, 8998-9002 (1988)), and restriction-site PCR, which uses universal primers to retrieve unknown nucleic acid sequence adjacent to a known locus (Sarkar, G. (1993) PCR Methods Applic., 2:318-322). "Inverse PCR" may also be used to amplify or to extend sequences using divergent primers based on a known region (Triglia, T. *et al.*, (1988) Nucleic Acids Res. 16:8186). Another method which may be used is "capture PCR", which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. *et al.*, (1991) PCR Methods Applic., 1:111-119). Another method which may be used to retrieve unknown sequences is that of Parker, J.D. *et al.*, (1991); Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and libraries, such as the PromoterFinder™ library (Clontech, Palo Alto, CA) to walk genomic DNA. This latter process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences that contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

In one embodiment of the invention, a nucleic acid molecule according to the invention may be used for chromosome localisation. In this technique, a nucleic acid molecule is specifically targeted to, and can hybridise with, a particular location on an individual human chromosome.

The mapping of relevant sequences to chromosomes is an important step in the confirmatory correlation of those sequences with the gene-associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationships between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes). This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localised by genetic linkage to a particular genomic region, any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleic acid molecule may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

Nucleic acid molecules of the present invention are also valuable for tissue localisation. Such techniques facilitate the determination of expression patterns of the polypeptide in tissues by detection of the mRNAs that encode them. These techniques include *in situ* hybridisation techniques and nucleotide amplification techniques, such as PCR. Results from these studies provide an indication of the normal functions of the polypeptide in the organism, as well as highlighting the involvement of a particular gene in a disease state or abnormal physiological condition.

In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by a mutant gene provide valuable insights into the role of mutant polypeptides in disease. Such inappropriate expression may be of a temporal, spatial or quantitative nature.

Vectors

The nucleic acid molecules of the present invention may be incorporated into vectors for cloning (for example, pBluescript made by Stratagene) or expression purposes. Vectors containing a nucleic acid molecule explicitly identified herein (or a variant thereof) form another aspect of this invention. The nucleic acid molecule may be inserted into an appropriate vector by any variety of well known techniques such as those described in Sambrook *et al.* [supra]. Generally, the encoding gene can be placed under the control of a control element such

as a promoter, ribosome binding site or operator, so that the DNA sequence encoding the desired polypeptide is transcribed into RNA in the transformed host cell.

Vectors may be derived from various sources including, but not limited to bacterial plasmids, bacteriophage, transposons, yeast episomes, insertion elements, yeast chromosomal elements, viruses for example, baculoviruses and SV40 (simian virus), vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, or combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, including cosmids and phagemids. Human, bacterial and yeast artificial chromosomes (HACs, BACs and YACs respectively) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid.

Examples of retroviruses include but are not limited to: murine leukaemia virus (MLV), human immunodeficiency virus (HIV), equine infectious anaemia virus (EIAV), mouse mammary tumour virus (MMTV), Rous sarcoma virus (RSV), Fujinami sarcoma virus (FuSV), Moloney murine leukaemia virus (Mo-MLV), FBR murine osteosarcoma virus (FBR MSV), Moloney murine sarcoma virus (Mo-MSV), Abelson murine leukaemia virus (A-MLV), Avian myelocytomatosis virus-29 (MC29), and Avian erythroblastosis virus (AEV). A detailed list of retroviruses may be found in Coffin et al ("Retroviruses" 1997 Cold Spring Harbour Laboratory Press Eds: JM Coffin, SM Hughes, HE Varmus pp 758-763).

Lentiviruses can be divided into primate and non-primate groups. Examples of primate lentiviruses include but are not limited to: the human immunodeficiency virus (HIV), the causative agent of human auto-immunodeficiency syndrome (AIDS), and the simian immunodeficiency virus (SIV). The non-primate lentiviral group includes the prototype "slow virus" visna/maedi virus (VMV), as well as the related caprine arthritis-encephalitis virus (CAEV), equine infectious anaemia virus (EIAV) and the more recently described feline immunodeficiency virus (FIV) and bovine immunodeficiency virus (BIV).

A distinction between the lentivirus family and other types of retroviruses is that lentiviruses have the capability to infect both dividing and non-dividing cells (Lewis et al 1992 EMBO. J 11: 3053-3058; Lewis and Emerman 1994 J. Virol. 68: 510-516). In contrast, other retroviruses - such as MLV - are unable to infect non-dividing cells such as those that make up, for example, muscle, brain, lung and liver tissue.

A vector may be configured as a split-intron vector. A split intron vector is described in PCT patent applications WO 99/15683 and WO 99/15684.

If the features of adenoviruses are combined with the genetic stability of retroviruses/lentiviruses then essentially the adenovirus can be used to transduce target cells to become transient retroviral producer cells that could stably infect neighbouring cells. Such retroviral producer cells engineered to express an antigen of the present invention can be implanted in organisms such as animals or humans for use in the treatment of angiogenesis and/or cancer.

Poxvirus vectors are also suitable for use in accordance with the present invention. Pox viruses are engineered for recombinant gene expression and for the use as recombinant live vaccines. This entails the use of recombinant techniques to introduce nucleic acids encoding foreign antigens into the genome of the pox virus. If the nucleic acid is integrated at a site in the viral DNA which is non-essential for the life cycle of the virus, it is possible for the newly produced recombinant pox virus to be infectious, that is to say to infect foreign cells and thus to express the integrated DNA sequence. The recombinant pox virus prepared in this way can be used as live vaccines for the prophylaxis and/or treatment of pathologic and infectious disease.

For vaccine delivery, preferred vectors are vaccinia virus vectors such as MVA or NYVAC. Most preferred is the vaccinia strain modified virus ankara (MVA) or a strain derived therefrom. Alternatives to vaccinia vectors include avipox vectors such as fowlpox or canarypox known as ALVAC and strains derived therefrom which can infect and express recombinant proteins in human cells but are unable to replicate.

Bacterial vectors may be also used, such as salmonella, listeria and mycobacteria.

Vectors containing the relevant nucleotide sequence may enter the host cell by a variety of methods well known in the art and described in many standard laboratory manuals (such as Sambrook *et al.*, [supra], Ausubel *et al.*, [supra], Davis *et al.*, Basic Methods in Molecular Biology (1986)). Methods include calcium phosphate transfection, cationic lipid-mediated transfection, DEAE-dextran mediated transfection, electroporation, microinjection, scrape loading, transduction, and ballistic introduction or infection.

Host cells

The choice of host cells is often dependent on the vector type used as a carrier for the nucleic acid molecule of the present invention. Bacteria and other microorganisms are particularly suitable hosts for plasmids, cosmids and expression vectors generally (for example, vectors derived from the pBR322 plasmid), yeast are suitable hosts for yeast expression vectors, insect cell systems are suitable host for virus expression vectors (for example, baculovirus) and plant cells are suitable hosts for vectors such as the cauliflower mosaic virus (CaMV) and tobacco

mosaic virus (TMV). Other expression systems include using animal cells (for example, with the LentiVectors™, Oxford BioMedica) as a host cell or even using cell-free translating systems. Some vectors, such as "shuttle vectors" may be maintained in a variety of host cells. An example of such a vector would be pEG 202 and other yeast two-hybrid vectors which can be maintained in both yeast and bacterial cells (see Ausubel *et al.*, [supra] and Gyuris, J., Cell, 75, 791-803).

Examples of suitable bacterial hosts include *Streptococci, Staphylococci, Escherichia coli, Streptomyces* and *Bacillus subtilis* cells. Yeast and fungal hosts *include Saccharomyces cerevisiae* and *Aspergillus* cells. Mammalian cell hosts include many immortalised cell lines available from the American Type Culture Collection (ATCC) such as CHO (Chinese Hamster Ovary) cells, HeLa cells, BHK (baby hamster kidney) cells, monkey kidney cells, C127, 3T3, BHK, HEK 293, Bowes melanoma and human hepatocellular carcinoma (for example, Hep G2) cells. Insect host cells that are used for baculovirus expression include Drosophila S2 and Spodoptera Sf9 cells. Plant host cells include most plants from which protoplasts be isolated and cultured to give whole regenerated plants. Practically, all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugar cane, sugar beet, cotton, fruit and other trees, legumes and vegetables.

Expression systems

Also included in present invention are expression vectors that comprise a nucleic acid molecule as described above. Expression vectors and host cells are preferably chosen to give long term, high yield production and stable expression of the recombinant polypeptide and its variants.

Expression of a polypeptide can be effected by cloning an encoding nucleic acid molecule into a suitable expression vector and inserting this vector into a suitable host cell. The positioning and orientation of the nucleic acid molecule insert with respect to the regulatory sequences of the vector is important to ensure that the coding sequence is properly transcribed and translated. Alternatively, control and other regulatory sequences may be ligated onto the nucleic acid molecule of this invention prior to its insertion into the expression vector. In both cases, the sequence of the nucleic acid molecule may have to be adjusted in order to effect correct transcription and translation (for example, addition of nucleotides may be necessary to obtain the correct reading frame for translation of the polypeptide from its encoding nucleic acid molecule).

A nucleic acid molecule of the invention may comprise control sequences that encode signal peptides or leader sequences. These sequences may be useful in directing the translated polypeptide to a variety of locations within or outside the host cell, such as to the lumen of the endoplasmic reticulum, to the nucleus, to the periplasmic space, or into the extracellular environment. Such signals may be endogenous to the nucleic acid molecules of the invention, or may be a heterologous sequence. These leader or control sequences may be removed by the host during post-translational processing.

A nucleic acid molecule of the present invention may also comprise one or more regulatory sequences that allow for regulation of the expression of polypeptide relative to the growth of the host cell. Alternatively, these regulatory signals may be due to a heterologous sequence from the vector. Stimuli that these sequences respond to include those of a physical or chemical nature such as the presence or absence of regulatory compounds, changing temperatures or metabolic conditions. Regulatory sequences as described herein, are non-translated regions of sequence such as enhancers, promoters and the 5' and 3' untranslated regions of genes. Regulatory sequences interact with host cellular proteins that carry out translation and transcription. These regulatory sequences may vary in strength and specificity. Examples of regulatory sequences include those of constitutive and inducible promoters. In bacterial systems, an example of an inducible promoter is the hybrid *lacZ* promoter of the Bluescript phagemid (Stratagene, LaJolla, CA) or pSportlTM plasmid (Gibco BRL). The baculovirus polyhedrin promoter may be used in insect cells.

An example of a preferred expression system is the lentivirus expression system, for example, as described in International patent application WO98/17815.

Detection of uptake of vectors by the host organism

Various methods are known in the art to detect the uptake of a nucleic acid or vector molecule by a host cell and/or the subsequent successful expression of the encoded polypeptide (see for example Sambrook *et al.*, [supra]).

Vectors frequently have marker genes that can be easily assayed. Thus, vector uptake by a host cell can be readily detected by testing for the relevant phenotype. Markers include, but are not limited to those coding for antibiotic resistance, herbicide resistance or nutritional requirements. The gene encoding dihydrofolate reductase (DHFR) for example, confers resistance to methotrexate (Wigler, M. *et al.* (1980) PNAS 77:3567-70) and the gene *npt* confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. *et al*

(1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, examples of which will be clear to those of skill in the art.

Markers however, only indicate that a vector has been taken up by a host cell but does not distinguish between vectors that contain the desired nucleic acid molecule and those that do not. One method of detecting for the said nucleic acid molecule is to insert the relevant sequence at a position that will disrupt the transcription and translation of a marker gene. These cells can then be identified by the absence of a marker gene phenotype. Alternatively, a marker gene can be placed in tandem with a sequence encoding a polypeptide of the invention under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

More direct and definitive methods to detect the presence of the nucleic acid molecule of the present invention include DNA-DNA or DNA-RNA hybridisation with a probe comprising the relevant antisense molecule, as described above. More direct methods to detect polypeptide expression include protein bioassays for example, fluorescence activated cell sorting (FACS), immunoassay techniques such as ELISA or radioimmunoassays.

Alternative methods for detecting or quantitating the presence of the nucleic acid molecule or polypeptide of this invention include membrane, solution or chip-based technologies (see Hampton, R. *et al.*, (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN) and Maddox, D.E. *et al.*, (1983) J. Exp. Med, 158, 1211-1216).

Transgenic animals

In another embodiment of this invention, a nucleic acid molecule according to the invention may be used to create a transgenic animal, most commonly a rodent. The modification of the animal's genome may either be done locally, by modification of somatic cells or by germ line therapy to incorporate inheritable modifications. Such transgenic animals may be particularly useful in the generation of animal models for drug molecules effective as modulators of the polypeptides of the present invention.

Polypeptide purification

A polypeptide according to the invention may be recovered and purified from recombinant cell cultures by methods including, but not limited to cell lysis techniques, ammonium sulphate precipitation, ethanol precipitation, acid extraction, anion or cation chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity

chromatography, hydroxylapatite chromatography and lectin chromatography, high performance liquid chromatography (HPLC) or fast performance liquid chromatography (FPLC). The polypeptide may need refolding after purification or isolation and many well known techniques are available that will help regenerate an active polypeptide conformation.

Many expression vectors are commercially available that aid purification of the relevant polypeptide. These include vectors that join the sequence encoding the polypeptide to another expressed sequence creating a fused protein that is easier to purify. Ways in which these fused parts can facilitate purification of the polypeptide of this invention include fusions that can increase the solubility of the polypeptide, joining of metal chelating peptides (for example, histidine-tryptophan modules) that allow for purification with immobilised metals, joining of protein A domains which allow for purification with immobilised immunoglobulins and the joining of the domain that is utilised in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). Fusion of the polypeptide of this present invention with a secretion signal polypeptide may also aid purification. This is because the medium into which the fused polypeptide has been secreted can subsequently be used to recover and purify the expressed polypeptide.

If necessary, these extraneous polypeptides often comprise a cleavable linker sequence which allows the polypeptide to be isolated from the fusion. Cleavable linker sequences between the purification domain and the polypeptide of the invention include those specific for Factor Xa or for enterokinase (Invitrogen, San Diego, CA). One such expression vector provides for expression of a fusion protein containing the polypeptide of the invention fused to several histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification by IMAC (immobilised metal ion affinity chromatography as described in Porath, J. *et al.* (1992), Prot. Exp. Purif. 3: 263-281), while the thioredoxin or enterokinase cleavage site provides a means for purifying the polypeptide from the fusion protein. A discussion of vectors that contain fusion proteins is provided in Kroll, D.J. *et al.* (1993; DNA Cell Biol. 12:441-453).

Assays

Another aspect of this invention includes assays that may be carried out using a polypeptide or nucleic acid molecule according to the invention. Such assays may be for many uses including the development of drug candidates, for diagnostic purposes or for the gathering of information for therapeutics.

If the polypeptide is to be expressed for use in screening assays, generally it is preferred that it be produced at the surface of the host cell in which it is expressed. In this event, the host cells may be harvested prior to use in the screening assay, for example using techniques such as fluorescence activated cell sorting (FACS) or immunoaffinity techniques. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the expressed polypeptide. If polypeptide is produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

The polypeptide of the invention can be used to screen libraries of compounds in any of a variety of drug screening techniques. Such compounds may activate (agonise) or inhibit (antagonise) the level of expression of the gene or the activity of the polypeptide of the invention and form a further aspect of the present invention. Examples of suitable compounds are those which are effective to alter the expression of a natural gene which encodes a polypeptide of the invention or to regulate the activity of a polypeptide of the invention.

Agonist or antagonist compounds may be isolated from, for example, cells, cell-free preparations, chemical libraries or natural product mixtures. These agonists or antagonists may be natural or modified substrates, ligands, enzymes, receptors or structural or functional mimetics. For a suitable review of such screening techniques, see Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).

Potential agonists or antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to the polypeptide of the invention and thereby modulate its activity. In this fashion, binding of the polypeptide to normal cellular binding molecules may be potentiated or inhibited, such that the normal biological activity of the polypeptide is enhanced or prevented.

The polypeptide of the invention that is employed in such a screening technique may be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. In general, such screening procedures may involve using appropriate cells or cell membranes that express the polypeptide that are contacted with a test compound to observe binding, or stimulation or inhibition of a functional response. The functional response of the cells contacted with the test compound is then compared with control cells that were not contacted with the test compound. Such an assay may assess whether the test compound results in a signal generated by activation of the polypeptide, using an appropriate detection system. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist in the presence of the test compound is observed.

Alternatively, simple binding assays may be used, in which the adherence of a test compound to a surface bearing the polypeptide is detected by means of a label directly or indirectly associated with the test compound or in an assay involving competition with a labelled competitor. In another embodiment, competitive drug screening assays may be used, in which neutralising antibodies that are capable of binding the polypeptide specifically compete with a test compound for binding. In this manner, the antibodies can be used to detect the presence of any test compound that possesses specific binding affinity for the polypeptide.

Assays may also be designed to detect the effect of added test compounds on the production of mRNA encoding the polypeptide in cells. For example, an ELISA may be constructed that measures secreted or cell-associated levels of polypeptide using monoclonal or polyclonal antibodies by standard methods known in the art, and this can be used to search for compounds that may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues. The formation of binding complexes between the polypeptide and the compound being tested may then be measured.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the polypeptide of interest (see International patent application WO84/03564). In this method, large numbers of different small test compounds are synthesised on a solid substrate, which may then be reacted with the polypeptide of the invention and washed. One way of immobilising the polypeptide is to use non-neutralising antibodies. Bound polypeptide may then be detected using methods that are well known in the art. Purified polypeptide can also be coated directly onto plates for use in the aforementioned drug screening techniques.

A polypeptide according to the invention may be used to identify membrane-bound or soluble receptors, through standard receptor binding techniques that are known in the art, such as ligand binding and crosslinking assays in which the polypeptide is labelled with a radioactive isotope, is chemically modified, or is fused to a peptide sequence that facilitates its detection or purification, and incubated with a source of the putative receptor (for example, a composition of cells, cell membranes, cell supernatants, tissue extracts, or bodily fluids). The efficacy of binding may be measured using biophysical techniques such as surface plasmon resonance and spectroscopy. Binding assays may be used for the purification and cloning of the receptor, but may also identify agonists and antagonists of the polypeptide, that compete with the binding of the polypeptide to its receptor. Standard methods for conducting screening assays are well understood in the art.

A typical polypeptide-based assay might involve contacting the appropriate cell(s) or cell membrane(s) expressing the polypeptide with a test compound. In such assays, a polypeptide according to the invention may be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. Any response to the test compound, for example a binding response, a stimulation or inhibition of a functional response may then be compared with a control where the cell(s) or cell membrane(s) was/were not contacted with the test compound.

A binding response could be measured by testing for the adherence of a test compound to a surface bearing a polypeptide according to the invention. The test compound may aid polypeptide detection by being labelled, either directly or indirectly. Alternatively, the polypeptide itself may be labelled, for example, with a radioisotope, by chemical modification or as a fusion with a peptide or polypeptide sequence that will facilitate polypeptide detection. Alternatively, a binding response may be measured, for example, by performing a competition assay with a labelled competitor or *vice versa*. One example of such a technique is a competitive drug screening assay, where neutralising antibodies that are capable of specifically binding to the polypeptide compete with a test compound for binding. In this manner, the antibodies may be used to detect the presence of any test compound that possesses specific binding affinity for the polypeptide. Alternative binding assay methods are well known in the art and include, but are not limited to, cross-linking assays and filter binding assays. The efficacy of binding may be measured using biophysical techniques including surface plasmon resonance and spectroscopy.

High throughput screening is a type of assay which enables a large number of compounds to be searched for any significant binding activity to the polypeptide of interest (see patent application WO84/03564). This is particularly useful in drug screening. In this scenario, many different small test compounds are synthesised on to a solid substrate. The polypeptide is then introduced to this substrate and the whole apparatus washed. The polypeptide is then immobilised by, for example, using non-neutralising antibodies. Bound polypeptide may then be detected using methods that are well known in the art. Purified polypeptide may also be coated directly onto plates for use in the aforementioned drug screening techniques.

Assay methods that are also included within the terms of the present invention are those that involve the use of the genes and polypeptides of the invention in overexpression or ablation assays. Such assays involve the manipulation of levels of these genes/polypeptides in cells and assessment of the impact of this manipulation event on the physiology of the manipulated cells. For example, such experiments reveal details of signaling and metabolic pathways in which the particular genes/polypeptides are implicated, generate information regarding the identities of

polypeptides with which the studied polypeptides interact and provide clues as to methods by which related genes and proteins are regulated.

Another aspect of this invention provides for any screening kits that are based or developed from any of the aforementioned assays.

C. Pharmaceuticals

A further aspect of the invention provides a pharmaceutical composition suitable for modulating hypoxia and/or ischaemia, comprising a therapeutically-effective amount of a polypeptide, a nucleic acid molecule, vector or ligand as described above, in conjunction with a pharmaceutically-acceptable carrier. A composition containing a polypeptide, nucleic acid molecule, ligand or any other compound of this present invention (herein known as X) is considered to be "substantially free of impurities" (herein known as Y) when X makes up more than 85% mass per mass of the total [X+Y] mass. Preferably X comprises at least 90% of the total X+Y mass. More preferably X comprises at least 95%, 98% and most preferably 99% of the total X+Y mass.

Carriers

Carrier molecules may be genes, polypeptides, antibodies, liposomes or indeed any other agent provided that the carrier does not itself induce toxicity effects or cause the production of antibodies that are harmful to the individual receiving the pharmaceutical composition. Further examples of known carriers include polysaccharides, polylactic acids, polyglycolic acids and inactive virus particles. Carriers may also include pharmaceutically acceptable salts such as mineral acid salts (for example, hydrochlorides, hydrobromides, phosphates, sulphates) or the salts of organic acids (for example, acetates, propionates, malonates, benzoates). Pharmaceutically acceptable carriers may additionally contain liquids such as water, saline, glycerol, ethanol or auxiliary substances such as wetting or emulsifying agents, pH buffering substances and the like. Carriers may enable the pharmaceutical compositions to be formulated into tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions to aid intake by the patient. A thorough discussion of pharmaceutically acceptable carriers is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Dosage

The amount of component X in the composition should also be in therapeutically effective amounts. The phrase "therapeutically effective amounts" used herein refers to the amount of agent needed to treat, ameliorate, or prevent (for example, when used as a vaccine) a targeted disease or condition. An effective initial method to determine a "therapeutically effective

amount" may be by carrying out cell culture assays (for example, using neoplastic cells) or using animal models (for example, mice, rabbits, dogs or pigs). In addition to determining the appropriate concentration range for X to be therapeutically effective, animal models may also yield other relevant information such as preferable routes of administration that will give maximum effectiveness. Such information may be useful as a basis for patient administration. A "patient" as used in herein refers to the subject who is receiving treatment by administration of X. Preferably, the patient is human, but the term may also include animals.

The therapeutically-effective dosage will generally be dependent on the patient's status at the time of administration. Factors that may be taken into consideration when determining dosage include the severity of the disease state in the patient, the general health of the patient, the age, weight, gender, diet, time and frequency of administration, drug combinations, reaction sensitivities and the patient's tolerance or response to the therapy. The precise amount can be determined by routine experimentation but may ultimately lie with the judgement of the clinician. Generally, an effective dose will be from 0.01 mg/kg (mass of drug compared to mass of patient) to 50 mg/kg, preferably 0.05 mg/kg to 10 mg/kg. Compositions may be administered individually to a patient or may be administered in combination with other agents, drugs or hormones.

Routes of administration

Uptake of a pharmaceutical composition of the invention by a patient may be initiated by a variety of methods including, but not limited to enteral, intra-arterial, intrathecal, intramedullary, intramuscular, intranasal, intraperitoneal, intravaginal, intravenous, intraventricular, oral, rectal (for example, in the form of suppositories), subcutaneous, sublingual, transcutaneous applications (for example, see WO98/20734) or transdermal means.

Gene guns or hyposprays may also be used to administer the pharmaceutical compositions of the invention. Typically, the therapeutic compositions may be prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Direct delivery of the compositions can generally be accomplished by injection, subcutaneously, intraperitoneally, intravenously or intramuscularly, or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Inhibition of excessive activity

If a particular disease state is partially or completely caused by an inappropriate excess in the activity of a polypeptide according to the invention, several approaches are available for inhibiting this activity.

One approach comprises administering to a patient an inhibitor compound (antagonist) along with a pharmaceutically acceptable carrier in an amount effective to inhibit the function of the polypeptide, such as by blocking the binding of a ligand, substrate, enzyme, receptor, or by inhibiting a second signal, and thereby alleviating the abnormal condition. Such an antagonist molecule may, for example, be an antibody. Most preferably, such antibodies are chimeric and/or humanised to minimise their immunogenicity, as previously described.

In another approach, soluble forms of the polypeptide that retain binding affinity for the ligand, substrate, enzyme, receptor, in question, may be administered to the patient to compete with the biological activity of the endogenous polypeptide. Typically, the polypeptide may be administered in the form of a fragment that retains a portion that is relevant for the desired biological activity.

In an alternative approach, expression of the gene encoding the polypeptide can be inhibited using expression blocking techniques, such as by using antisense nucleic acid molecules (as described above), either internally generated or separately administered. Modifications of gene expression may be effected by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5' or regulatory regions (signal sequence, promoters, enhancers and introns) of the gene encoding the polypeptide. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. *et al.* (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY). The complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes. Such oligonucleotides may be administered or may be generated *in situ* from expression *in vivo*.

Gene silencing approaches may also be undertaken to down-regulate endogenous expression of a gene. RNA interference (RNAi) (Elbashir, SM et al., Nature 2001, 411, 494-498) is one method of sequence specific post-transcriptional gene silencing that may be employed. Short dsRNA oligonucleotides are synthesised *in vitro* and introduced into a cell. The sequence

specific binding of these dsRNA oligonucleotides triggers the degradation of target mRNA, reducing or ablating target protein expression.

In addition, expression of a polypeptide according to the invention may be prevented by using a ribozyme specific to the encoding mRNA sequence for the polypeptide. Ribozymes are catalytically active RNAs that can be natural or synthetic (see for example Usman, N, *et al.*, Curr. Opin. Struct. Biol (1996) 6(4), 527-33). Synthetic ribozymes can be designed to specifically cleave mRNAs at selected positions thereby preventing translation of the mRNAs into functional polypeptide. Ribozymes may be synthesised with a natural ribose phosphate backbone and natural bases, as normally found in RNA molecules. Alternatively the ribozymes may be synthesised with non-natural backbones, for example, 2'-O-methyl RNA, to provide protection from ribonuclease degradation and may contain modified bases.

Efficacy of the gene silencing approaches assessed above may be assessed through the measurement of polypeptide expression (for example, by Western blotting), and at the RNA level using TaqMan-based methodologies.

RNA molecules may be modified to increase their intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of non-traditional bases such as inosine, queosine and butosine, as well as acetyl-, methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine and uridine that are not as easily recognised by endogenous endonucleases.

Activation of a polypeptide activity

If a particular disease state is partially or completely due to a lowered level of biological activity from a polypeptide according to the invention, various methods may be used. An example of such a method includes administering a therapeutically effective amount of compound that can activate (i.e. an agonist) or cause increased expression of the polypeptide concerned. Administration of such a compound may be via any of the methods described previously.

Gene Therapy

Another aspect of the present invention provides for gene therapy methods involving nucleic acid molecules identified herein. Gene therapy may be used to affect the endogenous production of the polypeptide of the present invention by relevant cells in a patient. For

example, gene therapy can be used permanently to treat the inappropriate production of a polypeptide by replacing a defective gene with the corrected therapeutic gene.

Treatment may be effected either *in vivo* or *ex vivo*. *Ex vivo* gene therapy generally involves the isolation and purification of the patient's cells, introduction of the therapeutic gene into the cells and finally, the introduction of the genetically-altered cells back into the patient. *In vivo* gene therapy does not require the isolation and purification of patient cells prior to the introduction of the therapeutic gene into the patient. Instead, the therapeutic gene can be packaged for delivery into the host. Gene delivery vehicles for *in vivo* gene therapy include, but are not limited to, non-viral vehicles such as liposomes, replication-deficient viruses (for example, adenovirus as described by Berkner, K.L., in Curr. Top. Microbiol. Immunol., 158, 39-66 (1992)) or adeno-associated virus (AAV) vectors as described by Muzyczka, N., in Curr. Top. Microbiol. Immunol., 158, 97-129 (1992) and U.S. Patent No. 5,252,479. Alternatively, "naked DNA" may be directly injected into the bloodstream or muscle tissue as a form of *in vivo* gene therapy.

One example of a strategy for gene therapy including a nucleic acid molecule of this present invention may be as follows. A nucleic acid molecule encoding a polypeptide of the invention is engineered for expression in a replication-defective retroviral vector. This expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding the polypeptide, such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a patient for engineering cells *in vivo* and expression of the polypeptide *in vivo* (see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in Human Molecular Genetics (1996), T Strachan and A P Read, BIOS Scientific Publishers Ltd).

Genetic delivery of antibodies that bind to polypeptides according to the invention may also be effected, for example, as described in International patent application WO98/55607.

Vaccines

A further embodiment of the present invention provides that the polypeptides or nucleic acid molecules identified may be used in the development of vaccines. Where the aforementioned polypeptide or nucleic acid molecule is a disease-causing agent, vaccine development can involve the raising of antibodies against such agents. Where the aforementioned polypeptide or nucleic acid molecule is one that is up-regulated, vaccine development can involve the raising of antibodies or T cells against such agents (as described in WO00/29428).

Vaccines according to the invention may either be prophylactic (i.e. prevents infection) or therapeutic (i.e. treats disease after infection). Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with pharmaceutically-acceptable carriers as described above. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, and other pathogens.

Vaccination processes may involve the use of heterologous vectors eg: prime with MVA and boost with DNA.

Since polypeptides may be broken down in the stomach, vaccines comprising polypeptides are preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions that may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient, and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents.

The vaccine formulations of the invention may be presented in unit-dose or multi-dose containers. For example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

The technology referred to as jet injection (see, for example, www.powderject.com) may also be useful in the formulation of vaccine compositions.

In accordance with this aspect of the present invention, polypeptides can be delivered by viral or non-viral techniques. Non-viral delivery systems include but are not limited to DNA transfection methods. Here, transfection includes a process using a non-viral vector to deliver a antigen gene to a target mammalian cell. Typical transfection methods include electroporation, nucleic acid biolistics, lipid-mediated transfection, compacted nucleic acid-mediated transfection, liposomes, immunoliposomes, lipofectin, cationic agent-mediated, cationic facial amphiphiles (CFAs) (Nature Biotechnology 1996 14; 556), multivalent cations such as spermine, cationic lipids or polylysine, 1, 2,-bis (oleoyloxy)-3-(trimethylammonio) propane (DOTAP)-cholesterol complexes (Wolff and Trubetskoy 1998 Nature Biotechnology 16: 421) and combinations thereof.

Viral delivery systems include but are not limited to adenovirus vectors, adeno-associated viral (AAV) vectors, herpes viral vectors, influenza, retroviral vectors, lentiviral vectors or baculoviral vectors, venezuelan equine encephalitis virus (VEE), poxviruses such as: canarypox virus (Taylor et al 1995 Vaccine 13:539-549), entomopox virus (Li Y et al 1998 XIIth International Poxvirus Symposium p144. Abstract), penguine pox (Standard et al. J Gen Virol. 1998 79:1637-46) alphavirus, and alphavirus based DNA vectors.

In addition to the use of polypeptide-based vaccines, this aspect of the invention includes the use of genetically-based vaccines, for example, those vaccines that are effective through eliciting the expression of a particular gene (either endogenous or exogenously derived) in a cell, so targeting this cell for destruction by the immune system of the host organism.

A number of suitable methods for vaccination and vaccine delivery systems are described in International patent application WO00/29428.

D. Diagnostics

Another aspect of the present invention provides for the use of a nucleic acid molecule identified herein as a diagnostic reagent.

For example, a nucleic acid molecule may be detected or isolated from a patient's tissue and used for diagnostic purposes. "Tissue" as defined herein refers to blood, urine, any matter obtained from a tissue biopsy or any matter obtained from an autopsy. Genomic DNA from the tissue sample may be used directly for detection of a hypoxia-related condition. Alternatively, the DNA may be amplified using methods such as polymerase chain reaction (PCR), the ligase chain reaction (LCR), strand displacement amplification (SDA), or other amplification techniques (see Saiki *et al.*, Nature, 324, 163-166 (1986); Bej, *et al.*, Crit. Rev. Biochem. Molec. Biol., 26, 301-334 (1991); Birkenmeyer *et al.*, J. Virol. Meth., 35, 117-126 (1991) and Brunt, J., Bio/Technology, 8, 291-294 (1990)). Such diagnostics are particularly useful for prenatal and even neonatal testing.

A method of diagnosis of disease using a polynucleotide may comprise assessing the level of expression of the natural gene and comparing the level of encoded polypeptide to a control level measured in a normal subject that does not suffer from the disease or physiological condition that is being tested. The diagnosis may comprise the following steps:

a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule of the invention and the probe;

b) contacting a control sample with said probe under the same conditions used in step a); and

c) detecting the presence of hybrid complexes in said samples;

wherein detection of differing levels of the hybrid complex in the patient sample compared to levels of the hybrid complex in the control sample is indicative of the dysfunction.

A further aspect of the invention comprises a diagnostic method comprising the steps of:

a) obtaining a tissue sample from a patient being tested for disease;

b) isolating a nucleic acid molecule according to the invention from said tissue sample; and

c) diagnosing the patient for disease by detecting the presence of a mutation in the nucleic acid molecule which is associated with disease.

To aid the detection of nucleic acid molecules in the above-described methods, an amplification step, such as PCR, may be included. An example of this includes detection of deletions or insertions indicative of the dysfunction by a change in the size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridising amplified DNA to labelled RNA of the invention or alternatively, labelled antisense DNA sequences of the invention. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by assessing differences in melting temperatures. The presence or absence of the mutation in the patient may be detected by contacting DNA with a nucleic acid probe that hybridises to the DNA under stringent conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an unhybridised portion of the nucleic acid probe strand at any portion corresponding to a mutation associated with disease; and detecting the presence or absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation in the corresponding portion of the DNA strand.

Point mutations and other sequence differences between the reference gene and "mutant" genes can be identified by other well-known techniques, such as direct DNA sequencing or single-strand conformational polymorphism, (see Orita *et al.*, Genomics, 5, 874-879 (1989)). For example, a sequencing primer may be used with double-stranded PCR product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabelled nucleotides or by automatic sequencing procedures with fluorescent-tags. Cloned DNA segments may also be used as probes to detect specific DNA segments. The sensitivity of this method is greatly enhanced

when combined with PCR. Further, point mutations and other sequence variations, such as polymorphisms, can be detected as described above, for example, through the use of allele-specific oligonucleotides for PCR amplification of sequences that differ by single nucleotides.

DNA sequence differences may also be detected by alterations in the electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (for example, Myers *et al.*, Science (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton *et al.*, PNAS. USA (1985) 85: 4397-4401).

In addition to conventional gel electrophoresis and DNA sequencing, mutations such as microdeletions, aneuploidies, translocations, inversions, can also be detected by *in situ* analysis (see, for example, Keller *et al.*, DNA Probes, 2nd Ed., Stockton Press, New York, N.Y., USA (1993)), that is, DNA or RNA sequences in cells can be analysed for mutations without need for their isolation and/or immobilisation onto a membrane. FISH is presently the most commonly applied method and numerous reviews of FISH have appeared (see, for example, Trachuck *et al.*, Science, 250, 559-562 (1990), and Trask *et al.*, Trends, Genet., 7, 149-154 (1991)).

Arrays

In another embodiment of the invention, an array of oligonucleotide probes comprising a nucleic acid molecule according to the invention can be constructed to conduct efficient screening of genetic variants, mutations and polymorphisms. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee *et al.*, Science (1996), Vol 274, pp 610-613).

In one embodiment, the array is prepared and used according to the methods described in WO95/11995 (Chee *et al*); Lockhart, D. J. *et al.* (1996) Nat. Biotech. 14: 1675-1680); and Schena, M. *et al.* (1996) PNAS 93: 10614-10619). Oligonucleotide pairs may range from two to over one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support. In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler *et al*). In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a

substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536 or 6144 oligonucleotides, or any other number between two and over one million which lends itself to the efficient use of commercially-available instrumentation.

Diagnostics using polypeptides or mRNA

In addition to the methods discussed above, diseases may be diagnosed by methods comprising determining, from a sample derived from a subject, an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

Assay techniques that can be used to determine levels of a polypeptide of the present invention in a sample derived from a host are well-known to those of skill in the art and are discussed in some detail above (including radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays). One example of this aspect of the invention provides a diagnostic method which comprises the steps of: (a) contacting a ligand as described above with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

Protocols such as ELISA, RIA, and FACS for measuring polypeptide levels may additionally provide a basis for diagnosing altered or abnormal levels of polypeptide expression. Normal or standard values for polypeptide expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably humans, with antibody to the polypeptide under conditions suitable for complex formation The amount of standard complex formation may be quantified by various methods, such as by photometric means.

Antibodies which specifically bind to a polypeptide of the invention may be used for the diagnosis of conditions or diseases characterised by expression of the polypeptide, or in assays to monitor patients being treated with the polypeptides, nucleic acid molecules, ligands and other compounds of the invention. Antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for the polypeptide include methods that utilise the antibody and a label to detect the polypeptide in human body fluids or extracts of cells or tissues. The antibodies may be used with or without

modification, and may be labelled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules known in the art may be used, several of which are described above.

Quantities of polypeptide expressed in subject, control and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease. Diagnostic assays may be used to distinguish between absence, presence, and excess expression of polypeptide and to monitor regulation of polypeptide levels during therapeutic intervention. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials or in monitoring the treatment of an individual patient.

Diagnostic kits

A diagnostic kit of the present invention may comprise:

(a) a nucleic acid molecule of the present invention;

(b) a polypeptide of the present invention; or

(c) a ligand of the present invention.

In one aspect of the invention, a diagnostic kit may comprise a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to the invention; a second container containing primers useful for amplifying the nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease. The kit may further comprise a third container holding an agent for digesting unhybridised RNA.

In an alternative aspect of the invention, a diagnostic kit may comprise an array of nucleic acid molecules, an array of antibody molecules, and/or an array of polypeptide molecules, as discussed in more detail above.

Such kits will be of use in diagnosing a disease or susceptibility to disease, particularly inflammation, oncology, or cardiovascular disease.

Various aspects and embodiments of the present invention will now be described in more detail by way of example, with particular reference to polypeptides regulated differentially under hypoxic conditions as opposed to normoxic conditions. It will be appreciated that modification of detail may be made without departing from the scope of the invention.

The method used in the following examples for the identification of hypoxia-induced genes and proteins utilised an approach herein termed "Smartomics". This method is a differential expression screening method for identifying genetic elements that are involved in a particular cellular process and is described in detail in a co-pending, co-owned International patent application PCT/GB01/00758 entitled "Differential Expression Screening Method". The method involves comparing:

(a) gene expression in a first cell of interest; and

(b) gene expression in a second cell of interest, which cell comprises altered levels, relative to physiological levels, of a biological molecule implicated in the cellular process, due to the introduction into the second cell of a heterologous nucleic acid directing expression of a polypeptide; and

identifying a genetic element whose expression differs, wherein gene expression in said first and second cell of interest is compared under at least two different environmental conditions of oxygen concentration.

Brief description of the Figures

Figure 1: Northern blots performed to confirm overexpression of HIF-1α and EPAS1 using adenoviral gene transfer in transduced macrophages. RNA loading was as follows: Lanes 1,2: Macrophages transduced with the adenovirus AdApt ires-GFP. Lanes 3,4: Macrophages transduced with the adenovirus AdApt HIF-1α–ires-GFP. Lanes 4,5: Macrophages transduced with the adenovirus AdApt EPAS1–ires-GFP. In lanes 1,3,5 the macrophages were maintained in normoxia (20% O_2). In lanes 2,4,6 the macrophages were maintained in hypoxia (0.1% O_2). Positions of bands from an RNA size ladder are indicated to the right of each blot in kilobases (kb). Hybridisation probes were complementary to the genes HIF-1α (A), EPAS1 (B) and 28s ribosomal RNA (C).

Figure 2: A scatter plot of two representative RNA samples analysed using Research Genetics GeneFilters. RNA from non-transduced macrophages in normoxia (Y-axis) or hypoxia (X-axis) was hybridised to two Research Genetics GeneFilters GF200 arrays. Analysis was output as normalised intensity for each gene on the array, with two values per gene corresponding to the signals from normoxia and hypoxia. These values were plotted as a scatter graph, with each dot representing a gene on the array. Genes expressed at similar levels between the RNA samples are located at the x=y line. In this representation an indication is apparent of the dynamic range of detection.

Figure 3: Analysis of Lactate Dehydrogenase A expression with Smartomics. In section A, thumbnail images of spots corresponding to the lactate dehydrogenase-A (LDH-A) gene are shown. Contrast levels were set at a level to allow optimal visualisation of this gene, but are at a constant setting throughout this figure. Each strip of 6 images corresponds to a discrete array position or experiment, over the range of RNA samples. Figures beneath individual spot images are ratios of the normalised intensity of that spot compared to the reference condition (gfp; 20%O_2). Array location: Identity of the spot as defined by Research Genetics. Clone: IMAGE identification. The histogram (section B) shows the average of the figures shown and error bars are standard deviation. gfp: cells transduced with AdApt ires-GFP. Hif-1a: Cells transduced with AdApt Hif-1α-ires-GFP. Epas1: Cells transduced with AdApt Epas1-ires-GFP.

Figure 4: Analysis of Glyceraldehyde 3-phosphate dehydrogenase expression with Smartomics. In section A, thumbnail images of spots corresponding to the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene are shown. Contrast levels were set at a level to allow optimal visualisation of this gene, but are at a constant setting throughout this figure. Each strip of 6 images corresponds to a discrete array position or experiment, over the range of RNA samples. Figures beneath individual spot images are ratios of the normalised intensity of that spot compared to the reference condition (gfp; 20%O_2). Array location: Identity of the spot as defined by Research Genetics. Clone: IMAGE identification. The histogram (section B) shows the average of the figures shown and error bars are standard deviation. gfp: cells transduced with AdApt ires-GFP. Hif-1a: Cells transduced with AdApt Hif-1α-ires-GFP. Epas1: Cells transduced with AdApt Epas1-ires-GFP.

Figure 5: Analysis of Platelet derived growth factor beta expression with Smartomics. In section A, thumbnail images of spots corresponding to the Platelet derived growth factor beta (PDGF Beta) gene are shown. Contrast levels were set at a level to allow optimal visualisation of this gene, but are at a constant setting throughout this figure. Each strip of 6 images corresponds to a discrete array position or experiment, over the range of RNA samples. Figures beneath individual spot images are ratios of the normalised intensity of that spot compared to the reference condition (gfp; 20%O_2). Array location: Identity of the spot as defined by Research Genetics. Clone: IMAGE identification. For this gene, it is noted that different IMAGE clones corresponding to the same gene are present. The histogram (section B) shows the average of the figures shown and error bars are standard deviation. gfp: cells transduced with AdApt ires-GFP. Hif-1a: Cells transduced with AdApt Hif-1α-ires-GFP. Epas1: Cells transduced with AdApt Epas1-ires-GFP.

Figure 6: Analysis of Monocyte Chemotactic Protein-1 expression with Smartomics. In section A, thumbnail images of spots corresponding to the Monocyte Chemotactic Protein-1 (MCP-1) gene are shown. Contrast levels were set at a level to allow optimal visualisation of this gene, but are at a constant setting throughout this figure. Each strip of 6 images corresponds to a separate experiment, over the range of RNA samples. Figures beneath individual spot images are ratios of the normalised intensity of that spot compared to the reference condition (gfp; 20%O_2). Array location: Identity of the spot as defined by Research Genetics. Clone: IMAGE identification. The histogram (section B) shows the average of the figures shown and error bars are standard deviation. gfp: cells transduced with AdApt ires-GFP. Hif-1a: Cells transduced with AdApt Hif-1α-ires-GFP. Epas1: Cells transduced with AdApt Epas1-ires-GFP.

Figure 7: Discovery of a novel gene (Hs.16335) using Smartomics. In section A, thumbnail images of spots corresponding to the EST from UniGene cluster Hs.16335 are shown. Contrast levels were set at a level to allow optimal visualisation of this gene, but are at a constant setting throughout this figure. For this gene, contrast levels are at maximum. Each strip of 6 images corresponds to a separate experiment, over the range of RNA samples. Figures beneath individual spot images are ratios of the normalised intensity of that spot compared to the reference condition (gfp; 20%O_2). Array location: Identity of the spot as defined by Research Genetics. Clone: IMAGE identification. The histogram (section B) shows the average of the figures shown and error bars are standard deviation. gfp: cells transduced with AdApt ires-GFP. Hif-1a: Cells transduced with AdApt Hif-1α-ires-GFP. Epas1: Cells transduced with AdApt Epas1-ires-GFP.

Figure 8: Virtual Northern blot hybridisation to validate discovery of Hs.16335 by Smartomics. A) Hybridisation probe = Hs.16335. B) Hybridisation probe = β actin. Lanes 1-6 are the RNA samples used in Figures 3-7, from cells transduced with adenovirus. Lanes 7-10 are from non-transduced macrophages with (lanes 9,10) or without (lanes 7,8) prior activation. Histograms show relative mRNA expression levels, from phosphorimager analysis, relating to the Northern blots positioned above. Figures are relative expression ratios compared to gfp (20% O_2).

Figure 9: Plasmid map for pSMART CMV-HIF.

Figure 10: Plasmid map for pSMART CMV-empty.

EXAMPLES

Example 1: The use of Smartomics for the identification of hypoxia-regulated genes in macrophages

The Smartomics method has been utilised herein to improve the discovery of genes activated or repressed in response to hypoxia in primary human macrophages. This involves augmenting the natural response to hypoxia, by experimentally introducing a key regulator of the hypoxia response, namely hypoxia inducible factor 1α (HIF-1α), into a population of primary human macrophages and comparing gene expression in these cells with that in control cells.

Overexpression of HIF-1α was done either in isolation or was done in combination with exposing the cells to hypoxia. This allowed the detection of resulting gene expression changes that would otherwise have not been detectable in response to hypoxia alone.

Although HIF-1α is well known to mediate responses to hypoxia, other transcription factors are also known or suspected to be involved. These include a protein called endothelial PAS domain protein 1 (EPAS1) or HIF-2α, which shares 48% sequence identity with HIF-1α ("Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells." Tian H, McKnight SL, Russell DW. *Genes Dev.* 1997 Jan 1;11(1):72-82.). Evidence suggests that EPAS1 is especially important in mediating the hypoxia-response in certain cell types, and it is clearly detectable in human macrophages, suggesting a role in this cell type (Griffiths *et al., Gene Ther.* 2000 Feb;7(3):255-62). In the light of this, the current example also utilises overexpression of EPAS1, as an independent means of improving discovery of hypoxia-responsive genes, to overexpression of HIF-1α. It also illustrates an embodiment of the invention, whereby differences in the response to HIF-1α or EPAS1 (or other mediators of the hypoxia response) may be identified, with the goal of identifying therapeutic target molecules more suitable for specific and efficient treatment of disease.

The introduction of foreign gene sequences (i.e. HIF-1α or EPAS1) to primary macrophages may be achieved by recombinant adenovirus. A commercially available system was used herein to produce adenoviral particles involving the adenoviral transfer vector AdApt, the adenoviral genome plasmid AdEasy and the packaging cell line Per-c6 (Introgene, Leiden, The Netherlands). The standard manufacturer's instructions were followed.

Three derivatives of the AdApt transfer vector have been prepared, named AdApt ires-GFP, AdApt HIF-1α–ires-GFP and AdApt EPAS1-ires-GFP. In these vectors, for convenience,

AdApt was modified such that inserted genes (i.e. HIF-1α or EPAS1) expressed from the powerful cytomegalovirus (CMV) promoter were linked to the green fluorescent protein (gfp) marker, by virtue of an internal ribosome entry site (ires). Therefore presence of green fluorescence provides a convenient indicator of viral expression of HIF-1α or EPAS1 in transduced mammalian cells.

Standard molecular biology methods were used to construct the derivatives of AdApt, which included reverse transcriptase PCR (RT-PCR), transfer of DNA fragments between plasmids by restriction digestion, agarose gel DNA fragment separation, "end repairing" double stranded DNA fragments with overhanging ends to produce flush blunt ends, and DNA ligation. Subcloning steps were confirmed by DNA sequencing. These techniques are well known in the art, but reference may be made in particular to Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (1989) and Ausubel et al., Short Protocols in Molecular Biology (1999) 4th Ed, John Wiley & Sons, Inc.

Briefly, AdApt ires-GFP was made by inserting the encephalomyocarditis virus EMCV ires followed by the green fluorescent protein gene (GFP), into the end-repaired HpaI restriction site of AdApt, immediately downstream of and in the same orientation as the CMV promoter. Both EMCV ires and gfp sequences are widely used and can be obtained from commonly available plasmids. SEQ ID NO:39 recites the exact nucleotide sequence of the joined ires-GFP which was inserted into the AdApt plasmid.

The plasmid AdApt HIF-1α–ires-GFP was derived from AdApt ires-GFP by inserting the protein coding sequence of human HIF-1α between the CMV promoter and the ires-GFP elements of AdApt ires-GFP. To do this, human HIF-1α cDNA was cloned by RT-PCR from human mRNA, and the sequence was verified by comparison to the published HIF-1α cDNA nucleotide sequence (Genbank accession U22431). The HIF-1α sequence was ligated as an end-repaired fragment into the end-repaired AgeI restriction site of AdApt ires-GFP [this is also the AgeI restriction site of the parental vector AdApt immediately downstream of the CMV promoter]. The exact DNA sequence containing HIF-1α that was inserted into AdApt ires-GFP is shown in SEQ ID NO: 40.

The plasmid AdApt EPAS1–ires-GFP was derived from AdApt ires-GFP by inserting the protein coding sequence of human EPAS1 between the CMV promoter and the ires-GFP elements of AdApt ires-GFP. To do this, human EPAS1 cDNA was cloned by reverse transcriptase PCR (RT-PCR) from human mRNA, and the sequence was verified by

comparison to the published EPAS1 cDNA nucleotide sequence (GenBank accession U81984). The EPAS1 sequence was ligated as an end-repaired fragment into the end-repaired AgeI restriction site of AdApt ires-GFP [this is also the AgeI restriction site of the parental vector AdApt immediately downstream of the CMV promoter]. The exact DNA sequence containing EPAS1 which was inserted into AdApt ires-GFP is shown in SEQ ID NO:41

The adenoviral transfer vectors AdApt HIF-1α–ires-GFP and AdApt EPAS1-ires-GFP, were verified prior to production of adenoviral particles, for their ability to drive expression of functionally active HIF-1α or EPAS1 protein from the CMV promoter in mammalian cells. This was achieved by transient transfection luciferase-reporter assays as described (Boast *et al.*, *Hum Gene Ther.* 1999 Sep 1;10(13):2197-208).

Using the aforementioned Introgene adenoviral system, caesium-banded, pure adenoviral particles were produced for each of the vectors AdApt ires-GFP, AdApt HIF-1α–ires-GFP and AdApt EPAS1-ires-GFP. Following the Introgene manual, adenoviral preparations were quantitated by spectrophotometry, yielding values of viral particles (VP) per milliliter.

To isolate human macrophage, monocytes were derived from peripheral blood of healthy human donors. 100ml bags of buffy coat from the Bristol Blood Transfusion Centre (Bristol, UK) were mixed with an equal volume of RPMI1640 medium (Sigma). This was layered on top of 10ml ficol-paque (Pharmacia) in 50ml centrifuge tubes and centrifuged for 25 min at 800 x g. The interphase layer was removed, washed in MACS buffer (phosphate buffered saline pH 7.2, 0.5% bovine serum albumin, 2mM EDTA) and resuspended at 80 microliter per 10n7 cells. To this, 20 microliter CD14 Microbeads (Miltenyi Biotec) were added, and the tube incubated at 4 degrees for 15 min. Following this, one wash was performed in MACS buffer at 400 x g and the cells were resuspended in 3 ml MACS buffer and separated on an LS+ MACS Separation Column (Miltenyi Biotec) positioned on a midi-MACS magnet (Miltenyi Biotec). The column was washed with 3 x 3ml MACS buffer. The column was removed from the magnet and cells were eluted in 5 ml MACS buffer using a syringe. Cells were washed in culture medium (AIM V (Sigma) supplemented with 2% human AB serum (Sigma), and resuspended at 2 x 10n5 cells per ml in the same medium and placed in large teflon-coated culture bags (Sud-Laborbedarf GmbH, 82131 Gauting, Germany) and transferred to a tissue culture incubator (37 degrees, 5% CO_2) for 7-10 days. During this period, monocytes spontaneously differentiate to macrophages. This is confirmed by examining cell morphology using phase contrast microscopy. Cells are removed from the bags by placing at 4 degrees for 30 min and emptying the contents.

The macrophages were washed and resuspended in DMEM (Gibco, Paisley, UK) supplemented with 4% fetal bovine serum (Sigma). $4x10^6$ cells were plated into individual 10cm Primeria (Falcon) tissue culture dishes in a total volume of 8 ml per plate, with $6x10^9$ adenoviral particles per ml. Following culture for 16 hr, during which the macrophages adhere to the plate and are infected by the adenoviral particles, the medium is removed and replaced by AIM V medium supplemented with 2% human AB serum. A further 24 hr period of culture is allowed prior to experimentation, to allow gene expression from the transduced adenovirus.

The above dosage of adenoviral particles was determined to be the minimum amount required to achieve transduction of the majority (over 80%) of the macrophage population, using green fluorescence as a marker of gene transfer. This was confirmed using a separate adenoviral construct containing the LacZ reporter gene. By selecting the minimum dose of virus, possible non-specific effects of viral transfer are minimised.

For experimentation with hypoxia, identical culture dishes were divided into two separate incubators: One at 37 degrees, 5% CO_2, 95% air (=Normoxia) and the other at 37 degrees, 5% CO_2, 94.9% Nitrogen, 0.1% Oxygen (=Hypoxia). After 8 hours culture under these conditions, the dishes were removed from the incubator, placed on a chilled platform, washed in cold PBS and total RNA was extracted using RNazol B (Tel-Test, Inc; distributed by Biogenesis Ltd) following the manufacturer's instructions.

The design of this experiment was to obtain six populations of cells (referred to for simplicity as "cell types"), differing only in their treatment with adenovirus and/ or hypoxia, as shown below:

"Cell Type"	Adenovirus	Expressed gene	Oxygen condition	
1	AdApt ires-GFP	none	Normoxia	(20% Oxygen)
2	AdApt ires-GFP	none	Hypoxia	(0.1% Oxygen)
3	AdApt HIF-1α–ires-GFP	HIF-1α	Normoxia	(20% Oxygen)
4	AdApt HIF-1α–ires-GFP	HIF-1α	Hypoxia	(0.1% Oxygen)
5	AdApt EPAS1–ires-GFP	EPAS1	Normoxia	(20% Oxygen)
6	AdApt EPAS1–ires-GFP	EPAS1	Hypoxia	(0.1% Oxygen)

Gene discovery can be implemented by comparing gene expression profiles between these "cell types". According to conventional methods available in the literature, one would make comparisons between cell types 2 and 1. By implementing the Smartomics method, several other possibilities are seen. Firstly, a comparison can be made between cell types 3 or 5 and cell type 1. Here, the stimulus of overexpressing key molecules involved in the hypoxia response may exceed the natural response the hypoxia, as seen for cell type 2. Secondly, in a preferred embodiment of the invention, a comparison can be made between cell types 4 or 6 and cell type 1. In this situation, the natural response to hypoxia is being augmented or boosted by overexpressing key molecules involved in the hypoxia response. It should be noted that the experimental design illustrated above uses a control adenovirus in place of untreated cells. By doing this, any non-specific effects of viral transduction should occur equally throughout the analysis, and will disappear.

Although efficient adenoviral gene transfer was indicated by green fluorescence in the transduced macrophages, Northern blotting was used to absolutely confirm overexpression of HIF-1α and EPAS1. RNA samples extracted from cell types 1-6 as described above were analysed by Northern blotting (Figure 1). The RNA samples (8ug total RNA per lane) were electrophoresed on a formaldehyde denaturing 1% agarose gel, then transferred to a nylon membrane (Hybond-N, Amersham, UK), and sequentially hybridised with ^{33}P-labelled DNA probes complementary in nucleotide sequence to HIF-1α (Figure 1a), EPAS1 (Figure 1b) or 28S ribosomal RNA (Figure 1c). The methodology used for Northern blotting, probe hybridisation under stringent conditions, and removal of probes between hybridisations, is well known in the art.

In Figure 1a, it can be seen that all lanes contain a faint band of approximately 4 kb, corresponding to the endogenous HIF-1α mRNA. In lanes 3,4, which contain RNA from cells transduced with AdApt HIF-1α–ires-GFP, a much stronger band of a similar size is observed, indicating successful overexpression of HIF-1α.

In Figure 1b, it can be seen that all lanes contain a very faint band of approximately 5 kb, corresponding to the endogenous EPAS1 mRNA. In lanes 5,6, which contain RNA from cells transduced with AdApt EPAS1–ires-GFP, a much stronger band at approximately 4 kb is observed, indicating successful overexpression of EPAS1. The difference in size of the endogenous and overexpressed EPAS1 is due to the long untranslated region of the endogenous gene, which is of no consequence.

In Figure 1c, it can be seen that 28S ribosomal RNA is detected in all lanes, indicating equal loading of RNA on the gel.

By phosphorimager quantitative analysis of Figures 1a and 1b it is apparent that overexpression levels of both HIF-1α and EPAS1 are approximately 80-fold over the endogenous levels. Adenoviral-directed mRNA overexpression of these genes is not further augmented by hypoxia. For example, in Figure 1a, the band intensity for lane 4 does not exceed that for lane 3. However at the protein and functional levels, hypoxia potentiates the action of the proteins encoded by these mRNAs (Semenza GL. *Annu Rev Cell Dev Biol.* 1999;15:551-78. "Regulation of mammalian O2 homeostasis by hypoxia-inducible factor 1").

Global mRNA expression profiles from the RNA samples isolated from the six "cell types" were obtained using Research Genetics Human GeneFilters Release 1 (GF200) (Research Genetics, Huntsville, AL). This method uses pre-made arrays of DNA complementary to 5,300 genes covering a range of levels of characterisation, including sequences which only match unannotated ESTs or cDNA sequences of unknown function.

The arrays are nylon in composition, and are spotted with DNA derived from specific IMAGE consortium cDNA clones (http://image.llnl.gov/image/). The arrays are hybridised to RNA samples which have been radioactively labelled with the isotope ^{33}P to measure the abundance of individual genes within the RNA samples. Multiple RNA samples are labelled and hybridised in parallel to separate copies of the array, and spot hybridisation signals are compared between the RNA samples.

Key issues in array-based mRNA expression analysis are sensitivity and reliability. Currently two other methods are available; glass microarrays and DNA chips, both of which utilise fluorescently labelled RNA (Bowtell DD. *Nat Genet.* 1999 Jan;21(1 Suppl):25-32. "Options available--from start to finish--for obtaining expression data by microarray."). Although these methods are often believed to offer increased sensitivity over Nylon-based methods, this belief lacks definitive proof. To the contrary, a careful comparison of the three approaches shows that for similar amounts of unamplified RNA, the nylon-based radioactive method is superior (Bertucci F, Bernard K, Loriod B, Chang YC, Granjeaud S, Birnbaum D, Nguyen C, Peck K, Jordan BR. *Hum Mol Genet.* 1999 Sep;8(9):1715-22. "Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples."). The microarray and DNA chip methods require much larger amounts of RNA which are often not easily obtained from primary cells, or complicated amplification methods, which are liable to introduce error.

To demonstrate the sensitivity of the array-based gene expression method used in the current exemplification of Smartomics, a scatter plot of two representative RNA samples analysed in our laboratory using Research Genetics GeneFilters, demonstrates a range of detection approaching 4-logs (Figure 2). By comparison, arguably the most sophisticated array-based method, the DNA chip, is quoted as having a range of detection of 3-logs (Affymetrix).

Therefore, it is reasonable to assume that the improvements afforded by Smartomics regarding sensitivity issues, as illustrated by the current exemplification, could not easily be obtained by utilising an alternative array-based method. In any case, any potentially superior array methodology could be further improved by utilising the Smartomics invention described here.

RNA extracted from the 6 "cell types" as described above, was radioactively labelled and hybridised to separate copies of the Research Genetics Human GeneFilter GF200 (experiment #1). Methods provided by the manufacturer were followed (http://www.resgen.com/products/GF200_protocol.php3). Images of hybridised arrays were obtained using a Molecular Dynamics Storm phosphorimager. RNA was then stripped from the arrays, following the aforementioned protocol.

To ensure reproducibility, this procedure was repeated with the same RNA samples (experiment #2). The entire data set was then imported and analysed using Research Genetics Pathways 3.0 software, as explained in the Pathways 3.0 manual. Key aspects of the current analysis are summarised below:

Project Tree set-up

"Condition Pairs" mode was used to simultaneously analyse multiple experiments. "Condition" means several arrays hybridised to similar RNA samples, derived from the same "cell type".

Condition	"Cell Type"	Adenovirus	Oxygen	Experiment #
1	1	AdApt ires-GFP	Normoxia	1
1	1	AdApt ires-GFP	Normoxia	2
2	2	AdApt ires-GFP	Hypoxia	1
2	2	AdApt ires-GFP	Hypoxia	2
3	3	AdApt HIF-1α–ires-GFP	Normoxia	1
3	3	AdApt HIF-1α–ires-GFP	Normoxia	2
4	4	AdApt HIF-1α–ires-GFP	Hypoxia	1
4	4	AdApt HIF-1α–ires-GFP	Hypoxia	2
5	5	AdApt EPAS1–ires-GFP	Normoxia	1
5	5	AdApt EPAS1–ires-GFP	Normoxia	2
6	6	AdApt EPAS1–ires-GFP	Hypoxia	1
6	6	AdApt EPAS1–ires-GFP	Hypoxia	2

Normalisation set-up

The "all data points" option and Y. Chen algorithm with default settings were selected, as explained in the Pathways 3.0 manual. The two experiments were treated as separate normalisation groups, such that global differences between hybridisation signals from different arrays from the same experiment were corrected.

Comparison analysis

Pair-wise comparisons were made between condition 2 and condition 1
condition 3 and condition 1
condition 4 and condition 1
condition 5 and condition 1
condition 6 and condition 1

In other words, pair-wise comparisons were made using condition 1 (i.e. cell type 1) as the reference condition. This corresponds to cells transduced with the control adenovirus AdApt ires-GFP and placed under normal oxygen concentration (normoxia). Comparisons are made in

this way for all genes present on the Research Genetics GF200 array. By comparing conditions, the analysis considers data from both experiments #1 and #2.

Filter settings

Filtering was then done to select genes with expression ratios of above 2.0 for at least one of the five pair-wise comparisons detailed above. Genes with low signal intensities for all of the six conditions were automatically eliminated, using an Intensity II filter of min 0.2, max 1000. Genes that did not respond in a reproducible way in experiment #1 and #2, were automatically eliminated using the Students t-test filter (90% confidence level).

Results were output as expression profiles of individual genes, showing normalised signal intensity and expression ratio. A key advantage of analysis in Pathways 3.0 is that high magnification thumbnail images of individual spots are displayed. This allows visual verification that the area being measured truly covers the region containing the hybridised array spot, and that the spot is real and not a background artefact.

Minor differences between quantitative data and corresponding thumbnail images are sometimes seen even though the sampled area is clearly the *bona fide* array spot. For example, by eye there might seem to be a small difference between two spots, though the quantitative analysis might suggest a larger difference. It should be noted that thumbnail images are not normalised to compensate for global differences, and are limited in image quality. Greyscale images are inherently limited in their capacity to depict quantitative differences in intensity. Digital images generated by the Storm phosphorimager cover a linear dynamic range of 100,000 for a single pixel, whereas printed images can only be depicted as 256 shades of grey.

Results for three representative known hypoxia-regulated genes

As demonstration that overexpression of HIF-1α or EPAS1 together with hypoxia exposure is superior to using non-transduced hypoxic cells, in terms of discovering *bona fide* hypoxia-regulated genes, results are shown for genes which are already known in the art to be regulated in hypoxia.

Three genes have been selected which are represented as double spots on the Research Genetics GF200 array. Therefore, because the whole experiment was repeated, a total of four repeat comparisons are possible for these genes.

The lactate dehydrogenase A (LDH-A) gene is known in the art to be activated by hypoxia (Webster KA. *Mol Cell Biochem.* 1987 Sep;77(1):19-28. "Regulation of glycolytic enzyme RNA transcriptional rates by oxygen availability in skeletal muscle cells."). In Figure 3, it can

be seen that in response to hypoxia alone (gfp 0.1% O_2) there is on average a 2.24-fold increase in mRNA expression compared to normoxia (gfp 20% O_2).

By overexpressing HIF-1α there is on average a 3.39-fold increase in LDH-A expression, providing a significant improvement over the natural response (Figure 3; HIF-1α 20% O_2). By utilising a preferred embodiment of the Smartomics method, and simultaneously overexpressing HIF-1α in the presence of hypoxia, the average response of LDH-A is elevated further to 4.50-fold (Figure 3; HIF-1α 0.1% O_2).

In the prior art it has been established that HIF-1α is responsible for mediating the hypoxia-induced activation of LDH-A (Iyer NV, Kotch LE, Agani F, Leung SW, Laughner E, Wenger RH, Gassmann M, Gearhart JD, Lawler AM, Yu AY, Semenza GL. *Genes Dev.* 1998 Jan 15;12(2):149-62 "Cellular and developmental control of O2 homeostasis by hypoxia-inducible factor 1 alpha."). However it has never been envisaged or demonstrated that overexpression of HIF-1α in a stable manner using viral gene transfer techniques, both with or without simultaneous hypoxia, causes secondary changes in gene expression which are markedly greater than the natural hypoxia response. The response to hypoxia of LDH-A is also improved by overexpressing EPAS1 (Figure 3; EPAS1), though this is less dramatic than overexpressing HIF-1α.

Like LDH-A, the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene is known in the art to be activated by hypoxia (Webster KA. *Mol Cell Biochem.* 1987 Sep;77(1):19-28. "Regulation of glycolytic enzyme RNA transcriptional rates by oxygen availability in skeletal muscle cells."). In Figure 4, it can be seen that in response to hypoxia alone (gfp 0.1% O_2) there is on average a 1.52-fold increase in mRNA expression compared to normoxia.

By overexpressing HIF-1α there is on average a 3.33-fold increase in GAPDH expression, providing a significant improvement over the natural response (Figure 4; HIF-1α 20% O_2). By utilising the full embodiment of the Smartomics method, and simultaneously overexpressing HIF-1α in the presence of hypoxia, the average response of GAPDH is elevated further to 4.57-fold (Figure 4; HIF-1α 0.1% O_2).

In the published literature, it has been established that HIF-1α is responsible for mediating the hypoxia-induced activation of GAPDH (Iyer NV, Kotch LE, Agani F, Leung SW, Laughner E, Wenger RH, Gassmann M, Gearhart JD, Lawler AM, Yu AY, Semenza GL. *Genes Dev.* 1998 Jan 15;12(2):149-62 "Cellular and developmental control of O2 homeostasis by hypoxia-inducible factor 1 alpha."). However in the art, it has never been envisaged or demonstrated

that overexpression of HIF-1α in a stable manner using viral gene transfer techniques, both with or without simultaneous hypoxia, causes secondary changes in gene expression which are markedly greater than the natural hypoxia response.

For GAPDH, it can be seen that overexpression of EPAS1 (Figure 4; EPAS1 20% O_2 and 0.1% O_2), has a significantly smaller effect than overexpressing HIF-1α. This demonstrates a separate embodiment of the Smartomics method, whereby genes are identified which respond selectively or preferentially to overexpression of EPAS1 or HIF-1α.

Platelet derived growth factor beta (PDGF β) is also known in the art to be activated by hypoxia (Kourembanas S, Hannan RL, Faller DV. *J Clin Invest.* 1990 Aug;86(2):670-4 "Oxygen tension regulates the expression of the platelet-derived growth factor-B chain gene in human endothelial cells."). In Figure 5, it can be seen that in response to hypoxia alone (gfp 0.1% O_2) there is on average a 2.14-fold increase in mRNA expression compared to normoxia.

By overexpressing EPAS1, there is on average a 9.28-fold increase in PDGF β expression (Figure 5; EPAS1 20% O_2), providing a large improvement over the natural response. In this case, the combination of hypoxia and EPAS1 overexpression does not exceed the response of EPAS1 overexpression alone, indicating saturation of the dose-response (Figure 5; EPAS1 0.1% O_2).

From Figure 5, it is clear that there is a striking specificity in the response of PDGF β to EPAS1 and HIF-1α, in the opposite manner observed for GAPDH. Overexpression of HIF-1α alone has no significant effect on PDGF β, whereas overexpression of EPAS1 produces large effects. This demonstrates a separate embodiment of the Smartomics method, whereby genes are identified which respond selectively or preferentially to overexpression of EPAS1 or HIF-1α.

The gene encoding monocyte chemotactic protein 1 (MCP-1) is known in the art to respond to hypoxia in a negative fashion, by decreasing mRNA expression (Negus RP, Turner L, Burke F, Balkwill FR. *J Leukoc Biol.* 1998 Jun;63(6):758-65. "Hypoxia down-regulates MCP-1 expression: implications for macrophage distribution in tumors"). In Figure 6 it can be seen that in response to hypoxia alone (gfp 0.1% O_2) there is on average a 0.407-fold change (i.e. a 2.46 fold decrease) in mRNA expression compared to normoxia.

By overexpressing HIF-1α, there is on average a 0.243-fold change (i.e. a 4.11-fold decrease) in MCP-1 expression, providing a significant improvement over the natural response (Figure 6; HIF-1α 20% O_2). By utilising a preferred embodiment of the Smartomics method, and

simultaneously overexpressing HIF-1α in the presence of hypoxia, the average response of MCP-1 is further improved to a 0.112-fold change (i.e. an 8.93-fold decrease) (Figure 6; HIF-1α 0.1% O_2). Even more pronounced improvements in the hypoxia-induced inhibition of MCP-1 expression are obtained by overexpressing EPAS1 (Figure 6; EPAS1 20% O_2 and 0.1% O_2). This demonstrates a use of Smartomics to improve the discovery of genes that are inhibited or repressed by disease signals.

The finding that overexpressing HIF-1α or EPAS1 potentiates hypoxia-induced gene repression, as exemplified by MCP-1, is totally without precedent in this field. The structure of both HIF-1α and EPAS1 proteins is that they contain transactivation domains but not known transcriptional repressor domains (Pugh CW, O'Rourke JF, Nagao M, Gleadle JM, Ratcliffe PJ. *J Biol Chem.* 1997 Apr 25;272(17):11205-14. "Activation of hypoxia-inducible factor-1; definition of regulatory domains within the alpha subunit.").

The results explained above relate to an array gene expression analysis, in which over 50 genes were identified as being regulated in hypoxia, from a total set of approximately 5300 genes on the array. By focusing on genes known in the art to be regulated in hypoxia, and showing how the Smartomics method can significantly enhance the response, an argument is provided that Smartomics would provide an improved method for the identification of novel *bona fide* hypoxia-regulated genes. In the current study, this can also be shown directly, for novel genes which were discovered using the Smartomics method, as presented below. Because expression changes arising from a conventional analysis are also covered in this analysis (i.e. hypoxia / normoxia comparisons without viral overexpression), the advantage of the Smartomics invention is clearly demonstrated.

Table 1 lists unannotated genes or ESTs which were identified in this analysis as being activated in response to viral-directed overexpression, but which would not have been identified from a hypoxia / normoxia comparison as done in the prior art. The final five columns of Table 1 show expression ratios compared to cells transduced with AdApt–ires-GFP in normoxia. The first of these five columns is the response without Smartomics, and in all cases shown here, the levels are below significance. The other four columns represent results obtained using the Smartomics method, and significant responses are seen here. In particular, in the final rows of this table, novel genes are identified which show large responses to EPAS1 overexpression.

Table 1: Novel genes identified in macrophages by Smartomics

	NUCLEOTIDE		PROTEIN		RATIO (compared to gfp N)				
Title	Seq ID	Accession	Seq ID	Accession	gfp H	hif N	hif H	epas N	epas H
PRO0518 hypothetical protein	2	R11658	1	AAF69617	0.89	1.11	0.97	3.81	3.89
hypothetical protein LOC51317	4	R02569	3	AAF64262	1.13	1.31	1.32	2.92	2.63
ESTs, Highly similar to A53770	6	R00332	5	BAB15101	1.71	1.41	1.58	6.79	6.45
ESTs, Moderately similar to AF119917_63 PRO2831	8	N68173	7	none	0.85	2.44	1.85	1.67	1.66
ESTs	10	H82330	9	none	1.06	1.11	0.90	1.88	2.79
ESTs	12	T97204	11	none	1.25	1.20	0.84	2.03	2.76
ESTs	14	R25464	13	none	0.96	1.51	1.41	2.15	3.01
ESTs	16	R25464	15	none	1.12	1.70	1.35	2.23	2.92
ESTs	18	R95132	17	none	0.91	1.38	1.06	2.32	2.79
ESTs, Weakly similar to A49134 Ig kappa chain V-I region	20	N80371	19	none	1.70	1.26	2.02	2.07	1.87
ESTs	22	R09498	21	none	1.06	1.73	1.53	1.94	2.18
ESTs	24	N74648	23	none	0.94	0.78	1.01	3.39	3.13
ESTs	26	T86016	25	none	1.42	1.73	1.59	3.78	3.65
ESTs	28	N99839	27	none	0.98	2.02	1.46	2.88	3.91
ESTs	30	R06745	29	none	1.00	2.17	1.77	3.00	2.59
ESTs	32	N64734	31	none	1.44	0.97	1.36	9.50	10.29
ESTs	34	T85201	33	none	0.87	1.18	1.06	14.99	14.71

Column 1 is the gene title as used in the UniGene database on 16 Feb 2001. Nucleotide and protein acessions are from the Genbank database. The final five columns show expression levels expressed as a ratio compared to cells transduced with AdApt ires-GFP in normoxia. gfp H: Expression in cells transduced with AdApt ires-GFP in hypoxia. Hif N: Expression in cells transduced with AdApt Hif-1α-ires-GFP in normoxia. Hif H: Expression in cells transduced with AdApt Hif-1α-ires-GFP in hypoxia. EPAS N: Expression in cells transduced with AdApt Epas1-ires-GFP in normoxia. EPAS H: Expression in cells transduced with AdApt Epas1-ires-GFP in hypoxia.

Figure 7 shows the expression profile of one of these genes, corresponding to an EST (GenBank accession N64734; IMAGE clone 293336). In the UniGene EST database (http://www.ncbi.nlm.nih.gov/UniGene/) this EST is currently clustered with only two other ESTs with accessions AI051607 (IMAGE 1674154) and T87161 (IMAGE 293336). The UniGene cluster number is Hs.16335, and it is totally unannotated in the database. Sequence analysis shows that this rare sequence is incomplete and lacks information on the protein coding sequence. In the Ensembl database of human genome project gene annotation (http://www.ensembl.org/) blast searches of predicted or confirmed cDNA sequences do not identify this EST. It is therefore apparent that from public domain information, the gene corresponding to EST IMAGE 293336, is a truly novel and unannotated gene.

In Figure 7, thumbnail array spot images are shown at maximal contrast, such that the background signal is apparent. It can be seen that in response to hypoxia alone (gfp 0.1% O_2) there is on average a 1.4-fold increase in mRNA expression compared to normoxia. However, this is not significant, because it is derived from widely different ratios from individual experiments (2.41 and 0.46). From the thumbnail images for gfp 20% O_2 and gfp 0.1% O_2 it is evident that expression of the genes under these conditions is below the detection threshold of the array-based method. However, when the Smartomics invention is used, and EPAS1 is overexpressed using viral gene transfer methods, a clearly detectable response in seen, with induction ratios of over 8-fold (Figure 7; EPAS1 20% O_2 or 0.1% O_2). The expression profile in Figure 7 also demonstrates a separate embodiment of Smartomics, for the identification of genes which respond selectively to HIF-1α or EPAS1.

To confirm the results presented in Figure 7, a more sensitive method was used to study expression of the gene corresponding to IMAGE clone 293336, namely virtual Northern blotting. It should be noted that this method would not have been suitable for the original discovery that IMAGE clone 293336 is induced by hypoxia, because virtual Northern blotting

and similar methods do not allow simultaneous screening of large numbers of genes. The technique is similar to conventional Northern blotting, with the exception that double stranded cDNA corresponding to the mRNA population of expressed genes is resolved by electrophoresis and blotted onto a nylon membrane. It relies on a method of cDNA synthesis which produces full length cDNA molecules, which is commercially available (SMART PCR cDNA Synthesis Kit; Clontech Laboratories Inc, Palo Alto, CA, USA).

The method for virtual Northern blotting was followed as described in the instruction manual for the SMART PCR cDNA Synthesis Kit. Briefly, 600ng cDNA was synthesised from the six RNA samples used for array hybridisation. An additional four RNA samples were also processed, derived from non-transduced macrophages cultured in normoxia and hypoxia (6 hours at 0.1% O_2) both with and without pre-treatment for 16 hours with 100 ng/ml Lipopolysaccharide (*E.coli* 026:B6 Sigma, UK) and 1000 u/ml human gamma interferon (Sigma, UK). This combination of factors causes macrophage activation, a process key to the physiological and pathophysiological actions of the macrophage. All 10 cDNA samples were resolved on an agarose gel, and alkali transfer onto Hybond N+ membrane (AmershamPharmacia, UK) was carried out according to the Hybond N+ instructions. Stringent hybridisations with ^{33}P-labelled cloned cDNA probes were performed as for standard Northern blot hybridisation, which is well known in the art. cDNA probes were radiolabelled using a commercially available kit (Prime-a-Gene, Promega, UK). The virtual Northern blot was hybridised first with the cDNA insert of IMAGE clone 1674154 from UniGene cluster Hs.16335 (Figure 8a). The blot was then stripped, by a high temperature / low salt wash, and was re-probed with the protein coding region of the human β-actin gene (Figure 8b).

From Figure 8a, it can be seen that the mRNA corresponding to Hs.16335 is detected as a doublet band of approximately 4.5 kb. This gene is strongly induced by adenoviral-directed overexpression of EPAS1 (lanes 5,6), consistent with the array data from Figure 7. The higher induction ratios in this non-array analysis are due to increased sensitivity afforded by the virtual Northern technique. Unlike the array data, expression of Hs.16335 is within the range of detection for all RNA samples. Importantly, hypoxia alone is seen to cause an induction ratio of approximately 60-fold (Figure 8a; lanes 2, 8). Therefore Hs.16335 is identified as a *bone fide* hypoxia-regulated gene, despite being beneath the detection level of an array screen in the absence of the Smartomics method used herein.

The results in Figure 8a also demonstrate a separate embodiment of the Smartomics method, whereby genes are identified which respond selectively or preferentially to overexpression of

EPAS1 or HIF-1α. Overexpression of HIF-1α causes an induction ratio of 18.9-fold (lane 3), whereas overexpression of EPAS1 causes a much larger induction ratio of 141-fold (lane 5).

In Figure 8a lane 9, it is shown that activation of macrophages by LPS and TNFα causes a 10.8-fold increase in expression of the gene corresponding to Hs.16335. Therefore this novel gene is possibly relevant to the inflammatory functions of macrophages.

In Figure 8b expression of the human β-actin gene is found to be roughly constant throughout this experiment, consistent with the differences in Figure 8a being due to specific changes in gene expression.

Rapid amplification of cDNA ends (RACE) has been performed to clone the full length version of the gene corresponding to Hs.16335, based on the size of the cDNA size on the virtual Northern blot. Sequencing and functional analysis of this gene will possibly lead to the identification of a new therapeutic target molecule. Crucial to this process was the initial use of the Smartomics invention.

Confirmation of targets using hybridisation assays

To confirm the findings presented in Example 1, the IMAGE clones identified from the Research Genetics Human GeneFilters have now been fabricated by the authors into an independently produced and verified gene array (referred to herein as the "custom gene array"), composed of PCR-amplified insert DNA. The methods used to produce this array are common in the art, but the key points are summarised below.

IMAGE clones were obtained from the UK MRC HGMP Resource Centre (Hinxton, Cambridge CB10 1SB, UK) and were re-isolated as individual colonies and sequenced to verify the correct identity of the clone. In the majority of cases, the same IMAGE clone identified from the Research Genetics Human GeneFilters was selected, but in some instances these clones were not available and alternatives were selected, corresponding to the same gene.

Additional genes, with well-defined roles in various disease processes relevant to hypoxia, were also represented on the array, as derived from IMAGE clones. It is well established in the literature that genes with similar functions are often co-regulated at the mRNA level, as determined by microarray data clustering methods (Iyer VR *et al, Science*. 1999 283(5398):83-7; Eisen MB *et al Proc Natl Acad Sci U S A*. 1998 95(25):14863-8). This allows associations to be made between genes of unknown function (as present in the current specification) to genes of well defined function, in order to add significance to the former.

Normalisation is a key issue in array analysis. The custom gene array is a single colour type array, and contains a selection of additional IMAGE clones corresponding to genes which were empirically determined not to be affected by hypoxia and which are highly expressed in a wide range of human tissues and cell types. During data analysis, spot intensities were divided by the mean of all the reference genes shown below, each of which was present in quadruplicate on each array.

Gene	IMAGE clone Acc.
FLJ11102 fis clone PLACE1005646	AA464704
matrix Gla protein	AA155913
guanine nucleotide binding protein alpha stimulating 1	R43581
DKFZp434A1319	W74725
cDNA FLJ23280 fis clone HEP07194	AA669443
beta actin	(in house clone)
EF1a-like protein	AI817566
ribosomal protein L37a	W91881

IMAGE clone plasmid miniprep DNA was prepared and PCR amplified with flanking vector primers of the sequences GTTTTCCCAGTCACGACGTTG and TGAGCGGATAACAATTTCACACAG. This was then purified and concentrated by ethanol precipitation, and the presence of a single band and DNA concentration were determined by agarose gel electrophoresis and by digital imaging methods.

Purified PCR product corresponding to all the clones (IMAGE and non-IMAGE) were normalised to 0.5 mg/ ml by dilution. Arrays were fabricated onto Hybond N+ (Amersham) membranes using a BioRobotics TAS arrayer (Biorobotics, Cambridge CB37LW, UK) with a 500 micron pin tool. Using 384-well source plates and a 2x2 arraying format this array was relatively low density, thereby eliminating problems of spot-to-spot signal bleed. Also the large pin size and high source plate DNA concentration improves the sensitivity of detection. Post-arraying denaturation/ neutralisation was essentially as described by Bertucci F *et al.*, 1999 (*Oncogene* 18: 3905-3912).

Total RNA was extracted from cells using RNeasy (Qiagen) and 7 micrograms RNA was labelled with 100 microCi 33P dCTP using 2 micrograms poly dT (10-20 mer) as primer in a reverse transcription reaction. First strand RNA was then degraded under alkaline contitions, and this was then neutralised with Tris HCl pH 8.0, and the labelled cDNA was purified using

BioRad BioSpin-6 chromatography columns. Pre-hybridisation was performed in 4 ml Research Genetics MicroHyb solution supplemented with 10micrograms poly dA (10-20 mer) and 10 micrograms Cot-1 DNA, at 45 degrees for 2-3 hours. The cDNA was then denatured by heating and added to the pre-hybridisation, which was continued for 18-20hr. Washing steps were done as follows: 2xSSC/ 1% SDS 2x20min at 50 degrees and 0.5xSSC/ 1% SDS 10min at 55 degrees. Arrays were exposed to Amersham Low Energy phosphor screens for 24hr and scanned using a phosphorimager at 50 micron resolution. Image analysis was done using ArrayVision software (Imaging Research Inc). Tab delimited data files were exported and a full analysis performed using GeneSpring software (Silicon Genetics).

Using the described methodology a dynamic range of detection of 4 logs and a sensitivity of at least 1 / 50,000 is obtained, as determined by spike doping titration experiments. Having several technical differences compared to the Research Genetics Human GeneFilters as used in the initial filing, data from the custom gene array is expected to be quantitatively different.

As explained earlier, hypoxia may be associated with several inflammatory conditions, including rheumatoid arthritis. Moreover, other diseases associated with hypoxia, including cancer and atherosclerosis are accepted as having important inflammatory aspects and hypoxia itself can modulate immune cell functions of the macrophage (Lewis JS *et al.*, 1999 *J.Leukocyte Biol.*. 66: 889-900). As cytokines are central to the regulation of immune processes, the utility of hypoxia-regulated genes regarding inflammatory conditions may be supported if those genes also respond to key cytokines. In a series of experiments, primary human macrophages, cultured as described, were treated with recombinant human cytokines at 100 ng/ml, and gene expression changes were determined. In the table below, expression values are shown corresponding to the fold change compared to untreated cells (all in normoxia). Values of less than 1.0 represent decreased gene expression.

Clone	Seq ID	IMAGE acc	IL-10 24hr	IL-13 24hr	IL-4 24hr	TGFβ 96hr	IL-12 48hr	IL-12 6hr	IL-15 6hr	IL-17 6hr	IL-6 6hr	IL-6 72hr	IL1β 24hr	IL1β 6hr	TNFα 96hr
p1M13	31	N64734	1.75	0.69	0.61	1.18	0.66	0.61	0.92	0.59	1.34	1.25	0.76	0.62	1.85
p1G6	7	N68173	0.68	1.23	0.94	1.69	1.24	2.28	0.73	0.75	1.24	1.09	0.83	0.69	1.03
p1K11	9	H82330	0.97	1.36	0.97	1.99	1.30	2.73	0.90	0.92	1.36	1.41	1.35	1.00	0.95
p1K6	11	T97204	0.56	1.03	0.92	2.99	1.10	2.10	0.75	0.91	1.26	1.29	0.72	0.68	1.01
p1L6	13	R25464	0.70	1.29	0.67	2.38	1.70	2.84	1.01	0.78	1.02	1.08	1.05	0.89	0.84
p1E24	17	R95132	0.58	0.95	0.88	1.50	1.50	2.22	0.88	1.04	0.94	1.11	1.07	0.82	0.83
p1L13	19	N80371	1.69	1.52	2.22	1.60	0.82	0.92	1.03	1.27	1.13	1.00	1.29	0.88	1.53
p1I9	21	R09498	0.68	0.98	0.76	2.46	1.35	2.18	0.95	1.08	1.07	0.85	1.08	0.77	0.91
p1L18	29	R06745	0.92	0.83	0.81	2.83	1.46	1.61	1.08	0.84	1.32	1.10	0.99	0.85	1.05
p1M14	33	T85201	0.79	0.83	0.84	1.73	1.25	1.74	0.88	0.70	1.04	0.78	0.71	0.71	0.78
p1M2	3	R02569	0.89	0.77	0.68	2.67	1.14	1.29	0.92	0.39	1.14	1.02	0.55	0.66	0.64
p1L5	1	R11658	0.69	0.63	0.69	1.49	0.63	0.94	1.06	0.86	0.57	0.59	0.53	0.68	0.78
p1I6	23	N74648	0.85	0.57	0.65	1.19	0.68	0.91	1.20	0.87	0.83	0.84	0.54	0.65	0.79

We present below that in additional experiments two of the genes in particular have expression profiles supporting a novel role in the response to inflammatory stimuli. These genes are encoded by IMAGE clone accession N80371 (Seq ID:19/20; clone ID:p1L13) and by IMAGE clone accession N64734 (Seq ID:31/32; clone ID:p1M13). The response of macrophages to a combination of lipopolysaccharide (LPS; E.coli 026:B6 100ng/ml from Sigma) and gamma interferon (IFNγ; 100ng/ml) induces particularly potent cell activation. Conversely, exposure to the cytokine IL-10 causes de-activation. To investigate the effects of these cytokines on macrophages in more detail, expression changes were determined in both hypoxia and normoxia, with and without these cytokine combinations for a period of 6 hours.

IMAGE clone accession N64734 (Seq ID:31/32; clone ID:p1M13) showed the following gene expression changes, compared with untreated cells:

A) in response to hypoxia, gene expression was increased 4.38-fold.

B) in response to LPS/IFNγ in normoxia, gene expression was not significantly changed

C) in response to LPS/IFNγ in hypoxia, gene expression was increased 3.33-fold.

D) in response to IL-10 in normoxia, gene expression was increased 2.16-fold

E) in response to IL-10 in hypoxia, gene expression was increased 11.4-fold

This profile indicates responsiveness to hypoxia or to IL-10 alone, and a synergistic effect when both stimuli are combined, but no response to LPS/IFNγ.

IMAGE clone accession N80371 (Seq ID:19/20; clone ID:p1L13) showed the following gene expression changes, compared with untreated cells:

A) in response to hypoxia, gene expression was increased 2.85-fold.

B) in response to LPS/IFNγ in normoxia, gene expression was not significantly changed

C) in response to LPS/IFNγ in hypoxia, gene expression was increased 3.00-fold.

D) in response to IL-10 in normoxia, gene expression was increased 1.31-fold

E) in response to IL-10 in hypoxia, gene expression was increased 3.11-fold

Superoxide radicals are frequently found at sites of inflammation and can be generated in the laboratory using a combination of Xanthine oxidase (10 mU/ml) and Xanthine (0.5 mM). Following treatment of macrophages for 18hr, IMAGE clone accession N64734 (Seq ID:31/32; clone ID:p1M13) was induced in expression by 3.55-fold, and IMAGE clone accession N80371 (Seq ID:19/20; clone ID:p1L13) was induced in expression by 2.0-fold. Additionally, IMAGE clone accession N64734 (Seq ID:31/32; clone ID:p1M13) was induced in expression in response to 0.5 mM H_2O_2 by 1.63-fold.

Example 2: Generation of codon-optimised EIAV vector expressing HIF1-α

This example describes the generation of an EIAV-derived vector, pSMART CMV-HIF in which expression of HIF-1α is driven from a CMV promoter located internally within the vector (Figure 9). A similar vector backbone could be used to achieve expression of other genes for the purposes of differential screening as described in this patent.

The starting point for construction of pSMART CMV-HIF was pONY4.0Z (WO 99/32646) and Mitophanous *et al.*, Gene Ther. 1999 Nov;6(11):1808-18. In the first step, plasmid pONY4.0Z was converted into pONY8.0Z by introducing mutations which 1) prevented expression of TAT by creating an 83nt deletion in exon 2 of tat, 2) prevented S2 ORF expression by a 51nt deletion, 3) prevented REV expression by deletion of a single base within exon 1 of rev, and 4) prevented expression of the N-terminal portion of gag by insertion of T residues within the first and second ATG codons of the gag region, thereby changing the sequence to ATTG from ATG. With respect to the wild type EIAV sequence (Accession No. U01866) these correspond to deletion of 1) nt 5234-5316 inclusive, 2) nt 5346-5396 inclusive, and 3) nt 5538. The insertion of T residues (4)) was after nt 526 and nt 543. These alterations

were carried out using techniques readily practicable to one skilled in the art. The resulting vector, pONY8.0Z expresses none of the EIAV accessory proteins or any of the EIAV gag protein.

In the next step, the β-galactosidase reporter gene present in pONY8.0Z was replaced by the enhanced green fluorescence protein (eGFP) reporter gene to create pONY8G. This was done by transferring the SacII –KpnI fragment corresponding to the GFP gene and flanking sequences from pONY2.13GFP (WO 99/32646) into pONY8.0Z cut with the same enzymes.

The presence of sequences termed the central polypurine tract and central termination sequence (cPPT/CTS) has been suggested to improve the efficiency of gene delivery by HIV-1 based vectors to non-dividing cells (Zennou *et al.*, Cell. 2000 Apr 14;101(2):173-85, Follenzi *et al.*, Nat Genet. 2000 Jun;25(2):217-22). The analogous *cis*-acting element of EIAV is located in the polymerase coding region and can be obtained as a functional element by using PCR amplification from any plasmid which contains the EIAV polymerase coding region (for example pONY3.1, WO 99/32646) as follows. The PCR product includes the central polypurine tract and the central termination sequence (CTS). The oligonucleotide primers used in the PCR reaction were:

EIAV cPPT POS: CAGGTTATT**CTAGA**GTCGACGCTCTCATTACTTGTAAC

EIAV cPPT NEG: CGAATGCGTT**CTAGA**GTCGACCATGTTCACCAGGGATTTTG

The recognition sequence for *Xba*I is shown in bold face and allows insertion into the pONY8G backbone. Before insertion of the cPPT/CTS PCR product prepared as described above, pONY8G was modified to remove the central termination sequence (CTS) which was already present in the pONY8G vector. This was achieved by subcloning the *Sal*I to *Sca*I fragment encompassing the CTS and RRE region from pONY8.0Z into pSP72, prepared for ligation by digestion with *Sal*I and *Eco*RV. The CTS region was then excised by digestion with *Kpn*I and *Ppu*MI, the overhanging ends 'blunted' by T4 DNA polymerase treatment and then the ends religated. The modified EIAV vector fragment was then excised using *Sal*I and *Nhe*I and ligated into pONY8G prepared for ligation by digestion with the same enzymes. This new EIAV vector was termed pONY8G del CTS. pONY8G del CTS has two *Xba*I sites which flank the CMV-GFP cassette and the PCR product representing the cPPT/CTS, after digestion with *Xba*I can be ligated into either site after partial digestion. Ligation into these sites results in plasmids with the cPPT/CTS element in either the positive or negative senses. Clones in which the cPPT/CTS was in the positive sense (functionally active) at either the 5' or 3'-position were termed pONY8G 5'POS del CTS and pONY8G 3'POS del CTS, respectively.

Another vector, termed pONY8Z 5'POS del CTS was also made following a similar strategy to that used to make pONY8G 5'POS del CTS. Accordingly, the CTS sequence present in pONY8.0Z was removed in the same way to make pONY8Z del CTS and the cPPT/CTS sequence was introduced into the unique *Xba*I site just upstream of the CMV promoter in pONY8Z del CTS.

The pSMART CMV-HIF vector plasmid was derived from pONY8G 5'POS del CTS by replacement of the coding region for eGFP with that of HIF-1α. This was achieved by digestion of the latter with *Sac*II and *Not*I, which flank the eGFP gene, and ligation to a *Sac*II-*Not*I fragment obtained from plasmid AdApt HIF-1α-ires-GFP. Construction of plasmid AdApt HIF-1α-ires-GFP is as described in Example 2 above.

An additional derivative of pONY8G 5'POS del CTS was also made in order to produce vector preparations which serve as 'negative controls' in transduction experiments. This vector termed, pSMART CMV-empty (Figure 10) was made by digestion of pONY8G 5'POS del CTS with *Bsm*BI and *Not*I, which flank the eGFP gene, followed by religation. On the basis of sequence analysis of the transcript driven by the internal promoter, only a 3 amino acid peptide is expected to be produced in cells transduced with this vector.

The EIAV vectors described above were produced by transient co-transfection of 293T human embryonic kidney cells with either vector plasmid, pONY3.1 (which expresses the EIAV gag/pol protein) and an envelope expression plasmid, pRV67 (which encodes the vesicular stomatitis virus protein G, VSV-G) using the calcium phosphate precipitation method.

Twenty four hours before transfection the 293T cells were seeded at 3.6×10^6 cells per 10cm dish in 10ml of DMEM supplemented with glutamine, non-essential amino acids and 10% foetal calf serum. Transfections were carried out in the late afternoon and the cells were incubated overnight prior to replacement of the medium with 6ml of fresh media supplemented with sodium butyrate (5mM). After 7 hours the medium was collected and 6ml of fresh unsupplemented media added to the cells. The collected medium was cleared by low speed centrifugation and then filtered through 0.4micron filters.

Vector particles were then concentrated by low speed centrifugation (6,000g, JLA10.500 rotor) overnight at 4°C and the supernatant poured off, leaving the pellet in the bottom of the tube. The following morning the remaining tissue culture fluid was harvested, cleared and filtered. It was then placed on top of the pellet previously collected and overnight centrifugation repeated. After this the supernatant was decanted and excess fluid was drained. Then the

pellet was resuspended in formulation buffer to 1/1000 of the volume of starting supernatant. Aliquots were then stored at –80°C.

Formulation buffer (100ml)

Tissue culture grade water	28.65ml
19.75mM Tris/HCl buffer pH 7.0	19.75ml of a 0.1M solution
40mg/ml lactose	26.6ml of a 150mg/ml solution
37.5mM sodium chloride	24.4ml of a 154mM solution
1mg/ml human serum albumin[a]	500µl of a 20% solution
5µl/ml protamine sulphate[b]	100µl of a 5mg/ml solution

[a]Human serum albumin (20%) (Albutein, Alpha therapeutics UK Ltd, Thetford, Norfolk).

[b]Protamine sulphate 5mg/ml (Prosulf, CP Pharmaceuticals, Wrexham, UK).

The sequence of pSMART CMV-HIF is presented in SEQ ID NO:42

The sequence of pSMART CMV-empty is presented in SEQ ID NO:43

Example 3: Gene identification in hippocampal neurones

As discussed above in Example 1, hypoxia is an important component of stroke (cerebral ischaemia). The Smartomics method has now been utilised to improve the discovery of genes activated or repressed in response to hypoxia in primary rat hippocampal neurones. This involves augmenting the natural response to hypoxia, by experimentally introducing a key regulator of the hypoxia response, namely hypoxia inducible factor 1α (HIF-1α). The overexpression of HIF-1α in combination with exposure of the cells to hypoxia has allowed the detection of gene expression changes which would not been detectable in response to overexpression of HIF-1α alone, or hypoxia alone.

Primary rat hippocampal neuron cultures were established according to standard procedures from embryonic rats (Dunnett SB, Bjorkland A (Eds.) 1992. Neural Transplantation, A Practical Approach. IRL Press). Briefly, timed-pregnant Wistar rats at eighteen days of gestation were anaesthetised with 0.7 ml isofluorane and killed by cervical dislocation. Pups were removed from the uterus and decapitated. Hippocampi were dissected and stored on ice in Hanks Buffered Saline Solution (HBSS) containing DNAse (0.05%) and glucose (2 mM) before incubation in trypsin (0.1%) plus DNAse (0.05%) for 5 minutes. After incubation, trypsin was inactivated by the addition of soybean trypsin inhibitor (SBTI, 0.1%) and the

solution gently triturated. Cells were pelleted by centrifugation (3000 rpm, 5 minutes) and the trypsin removed. Cells were then washed twice in HBSS containing SBTI and DNAse (0.05%), and re-pelleted before final suspension in Dulbecco's Modified Eagle's Medium (DMEM) containing foetal calf serum (10%), glutamine (2 mM), and gentamicin (0.1 mg.ml^{-1}). Cells (3×10^6 cells per dish) were plated out onto 60 mm dishes coated with poly-D-Lysine (50 $\mu g.ml^{-1}$) and fibronectin adhesion promoting peptide (10 $\mu g.ml^{-1}$). Cultures were placed into a humidified 37°C incubator containing 5% CO_2 and twelve hours after plating, 50% of the plating medium was replaced with Neurobasal Media (Brewer GJ, 1995. Serum-free B27/neurobasal medium supports differentiated growth of neurons from the striatum, substantia nigra, septum, cerebral cortex, cerebellum, and dentate gyrus. Journal of Neuroscience Research 42:674-83) supplemented with B27 and glutamine (2 mM). Cultures were fed every two days with supplemented neurobasal medium and were transduced on day 3 *in vitro*.

Transduction was carried out in supplemented neurobasal media containing polybrene (2 $\mu g.ml^{-1}$), in 0.5 volumes of the typical culture media volume. Five hours after the onset of transduction, the media volume was increased by a factor of 2, and was replaced 12 hours later. The viruses pSMART CMV-HIF (carrying the HIF-1α gene; see Example 3), pSMART CMV-empty (an empty genome used as a control; see Example 3) and pONY8Z 5'POS del CTS (containing the β-galactosidase gene) were produced in parallel according to methods detailed above. The pONY8Z 5'POS del CTS was used to calculate viral titer in D17 cells and in hippocampal neurons. Comparison of the RNA packaging signal by quantitative RT-PCR (Taqman) of the three viral preps, allowed the biological titers of pSMART CMV-HIF and pSMART CMV-empty viruses to be estimated relative to that pONY8Z 5'POS del CTS. All transductions were done using approximately equal multiplicity of infections (MOIs) for both viruses, and the MOI used in each experiment was at least ten.

Thirty-six hours after transduction, identical culture dishes were divided into two separate incubators, one at 37°C, 5% CO_2, 95% air (=Normoxia) and the other at 37°C, 5% CO_2, 94.9% Nitrogen, 0.1% Oxygen (=Hypoxia). After 6 hours culture under these conditions, the dishes were removed from the incubator, placed on a chilled platform, washed in cold PBS and total RNA was extracted using RNazol B (Tel-Test, Inc; distributed by Biogenesis Ltd) following the manufacturer's instructions.

The experiment yielded four samples, differing only in their treatment with lentivirus and/or hypoxia, as shown below:

Sample	Lentivirus	Expressed gene	Oxygen condition
1	pSMART CMV-empty	none	Normoxia
2	pSMART CMV-empty	none	Hypoxia
3	pSMART CMV-HIF	HIF-1α	Normoxia
4	pSMART CMV-HIF	HIF-1α	Hypoxia

Gene discovery can be implemented by comparing gene expression profiles between these samples. According to conventional methods published in the art, one would make comparisons between cell types 1 and 2. By implementing the Smartomics method, several other possibilities are seen. Firstly, a comparison can be made between cell types 1 and 3. Here, the stimulus of overexpressing key molecules involved in the hypoxia response may exceed the natural response to hypoxia, as seen for cell type 2. Secondly, a comparison can be made between cell types 1 and 4. In this situation the natural response to hypoxia is being augmented or boosted by overexpressing key molecules involved in the hypoxia response.

Global mRNA expression profiles from the RNA isolated from the four samples were obtained using the Research Genetics Rat GeneFilter GF300 (Research Genetics, Huntsville, AL). This method uses pre-made nylon arrays of DNA derived from I.M.A.G.E./LLNL cDNA clones containing the 3' ends of genes (http://image.llnl.gov/image/). The arrays include more than 5,000 genes covering a range of levels of characterisation, including sequences which are representative of unannotated ESTs or cDNA sequences of unknown function.

RNA extracted from the 4 samples described above, was radioactively labelled and hybridised to separate copies of the Research Genetics Rat GeneFilter GF300. Methods provided by the manufacturer were followed (http://www.resgen.com/products/GF200_protocol.php3) with the following modifications; RNAsin was added to the labelling reaction, and following labelling the mRNA/cDNA hybrid was denatured by incubation with 45mM EDTA/18mM NaOH at 65°C for 30 minutes.

Images of hybridised arrays were obtained using a Molecular Dynamics Storm phosphorimager. RNA was then stripped from the arrays, following the aforementioned protocol. To ensure reproducibility, this procedure was repeated with the same RNA samples. Both data sets were then imported and analysed using Research Genetics Pathways 3.0

software, as explained in the Pathways 3.0 manual. Key aspects of the current analysis are summarised below:

Project Tree set-up

"Condition Pairs" mode was used to simultaneously analyse multiple experiments. In this context a condition is equivalent to a sample (e.g. Sample 3, overexpression of HIF-1α in normoxia).

Normalisation set-up

Data point normalisation was selected, as explained in the Pathways 3.0 manual. This technique generates normalised intensities by dividing all sampled intensities by the mean sampled intensity of all clones (except the control points) on the array. The two experiments were treated as separate normalisation groups, such that global differences in hybridisation signals between different arrays within the same experiment were corrected for.

Comparison analysis

Condition 1 (i.e. Sample 1) corresponds to cells transduced with the control lentivirus and placed under normal oxygen concentrations (normoxia). This was used as the reference condition in pairwise comparisons with conditions 2, 3 and 4 (i.e. samples 2, 3 and 4). Comparisons were made in this way for all genes present on the Research Genetics GF300 array. By comparing conditions the analysis considers data from both experiments.

Results for four representative known HIF-1α/hypoxia-regulated genes

As demonstration that overexpression of HIF-1α in hypoxic cells is superior to using non-transduced hypoxic cells or overexpression of HIF-1α in normoxic cells, in terms of discovering *bona fide* hypoxia-regulated genes, results are shown below for genes which are already known in the art to be regulated by hypoxia and HIF-1α. Ratios are expressed as average ratios of normalised intensities.

Table 2. Response of known HIF-1α /hypoxia-regulated genes

	PROTEIN		NUCLEOTIDE		RATIO SAMPLE 1 (normoxia) vs		
TITLE	SEQ ID	ACCESSION	SEQ ID	ACCESSION	SAMPLE 2	SAMPLE 3	SAMPI 4
					(hypoxia)	(Hif+normoxia)	(Hif+hy xia)
Enolase 1, alpha		NP_036686		NM_012554	1.04	0.86	1.40
Glucose-transporter protein		AAA41248		M13979	1.41	0.78	2.14
Glyceraldehyde-3-phosphate dehydrogenase		AAA40814		M29341	1.13	1.42	1.67
Lactate dehydrogenase A		CAA26000		X01964	1.36	1.50	1.77

All four genes listed in Table 2 are known in the art to be regulated by hypoxia, and specifically by Hif-1α (Iyer *et al* (1998) Cellular and developmental control of O_2 homeostasis by hypoxia-inducible factor 1α. Genes Dev 12:149-162). In the case of Enolase 1, alpha, the response to hypoxia or overexpression of Hif-1α under normoxia is undetectable. It is only when Hif-1α is overexpressed under hypoxia that an increase in expression level relative to normoxia is detected. In the case of glucose-transporter protein the detectable response to hypoxia is increased by the overexpression of Hif-1α in hypoxia. In the case of both glyceraldehyde-3-phosphate dehydrogenase and Lactate dehydrogenase A the response to hypoxia is detectable, but it is increased by the overexpression of Hif-1α under normoxia, and even more so by the overexpression of Hif-1α under hypoxia.

Filter settings

Data filtering was then performed to reduce the data set and select genes with expression ratios of above 2.0 for at least one of the three pair-wise comparisons detailed above. Genes with low signal intensities in all four conditions were automatically eliminated, using an Intensity II filter minimum of 0.2. Genes which did not respond in a reproducible way in both experiments were automatically eliminated using the Students t-test filter (90% confidence level).

Results were output as expression profiles of individual genes, showing normalised signal intensity and expression ratio. A key advantage of analysis in Pathways 3.0 is that high magnification thumbnail images of individual spots from the original images are displayed.

This allows visual verification that the area being measured truly covers the region containing the hybridised array spot.

Annotation of known and novel genes

As demonstration that overexpression of HIF-1α in hypoxic cells is superior to using non-transduced hypoxic cells or overexpression of HIF-1α in normoxic cells, in terms of discovering novel hypoxia-regulated genes, results are shown below for a gene which is already known in the art to be regulated by hypoxia, but not by HIF-1α, and for an unannotated gene. Ratios are expressed as average ratios of normalised intensities.

Table 3. Response of novel HIF-1α regulated genes

	PROTEIN		NUCLEOTIDE		RATIO SAMPLE 1 (normoxia) vs		
TITLE	SEQ ID	ACCESSION	SEQ ID	ACCESSION	SAMPLE 2 (hypoxia)	SAMPLE 3 (Hif+ normoxia)	SAMPLE 4 (Hif+ hypoxia)
Metallothionein-I [a]	37	AAA41590	38	J00750	1.61	1.24	3.49
EST	35	none	36	AA901269	1.43	1.08	3.47

[a] representative metallothionein ESTs are spotted twice on the array, so the data is the average of two points

Metallothionein-I is known in the literature to be regulated by hypoxia (Murphy *et al* (1999) Activation of metallothionein gene expression by hypoxia involves metal response elements and metal transcription factor-1. Cancer Res 59(6):1315-22), but it is not known to be regulated by HIF-1α. The data in Table 3 show that the response to overexpression of HIF-1α in hypoxia greatly exceeds that of hypoxia alone or the overexpression of HIF-1α in normoxia. The EST (expressed sequence tag) is a completely unannotated DNA sequence. Similarly, the data in Table 3 show that the response to overexpression of HIF-1α in hypoxia greatly exceeds that of hypoxia alone or the overexpression of HIF-1α in normoxia.

This data demonstrates that the methods described above enable the further functional annotation of known genes and the functional annotation of completely unannotated novel genes with no known function.

Table 4: Summary of novel genes and regulation mechanisms discussed herein

		Nucleotide			Protein			RATIO (compared to gfp N)				
Old Title	New Title	Seq ID	Accession	New Accession	Seq ID	Accession	New Accession	gfp H	hif N	hif H	epas N	epas H
PRO0518 hypothetical protein	Splice variant of MEG3	2	R11658	AB032607	1	AAF69617	None	0.89	1.11	0.97	3.81	3.89
hypothetical protein LOC51317	Hypothetical protein KIAA1696	4	R02569	XM_051010	3	AAF64262	XP_051010	1.13	1.31	1.32	2.92	2.63
ESTs, Highly similar to A53770	EGLN3	6	R00332	NM_022073	5	BAB15101	NP_071356	1.71	1.41	1.58	6.79	6.45
ESTs, Moderately similar to AF119917_63 PRO2831	Splice variant of PHPS1-2	8	N68173	BE671816	7	None	XP_096898	0.85	2.44	1.85	1.67	1.66
ESTs	ESTs	10	H82330	N68173	9	None	None	1.06	1.11	0.9	1.88	2.79
ESTs	IL6R	12	T97204	NM_000565	11	None	NP_000556	1.25	1.2	0.84	2.03	2.76
ESTs	hypothetical protein XP_091988	14	R25464	XM_091988	13	None	XP_091988	0.96	1.51	1.41	2.15	3.01
ESTs	ESTs	18	R95132	R95132	17	None	None	0.91	1.38	1.06	2.32	2.79
ESTs, Weakly similar to A49134 Ig kappa chain V-I region	FRA2	20	N80371	BC022791	19	None	P15408	1.7	1.26	2.02	2.07	1.87
ESTs	DBI	22	R09498	None	21	None	XP_038526	1.06	1.73	1.53	1.94	2.18
ESTs	BACH2	24	N74648	AJ271878	23	None	CAC28130	0.94	0.78	1.01	3.39	3.13
ESTs	ESTs	26	T86016	T86016	25	None	None	1.42	1.73	1.59	3.78	3.65
ESTs	Aldehyde dehydrogenase 12	28	N99839	NM_022568	27	None	NP_072090	0.98	2.02	1.46	2.88	3.91
ESTs	ESTs	30	R06745	R06745	29	None	None	1	2.17	1.77	3	2.59
ESTs	ESTs	32	N64734	N64734	31	None	None	1.44	0.97	1.36	9.5	10.29
ESTs	ESTs	34	T85201	T85201	33	None	None	0.87	1.18	1.06	14.99	14.71
Metallothionein-I a	Metallothionein-I a	38	J00750	J00750	37	AAA41590	AAA41590	1.61	1.24	3.49		

Example 4: Discussion of relevance of individual clones

The nucleotide sequence splice variant represented of MEG3, otherwise known as Gtl2, is located within an imprinting cluster on human Chromosome 14q32 (Wylie *et al* 2000, Genome Research vol 10 pp1711-8). The MEG3 gene contains no substantial open reading frame and apparently does not encode a protein. The nucleotide sequence of MEG3 is represented in the public sequence databases by the accession AB032607 and is described herein as SEQ ID No. 2. The Est R11658 represents a novel exon of the MEG3 gene of which there are many within the imprinting cluster. This maternally expressed imprinted gene may function as an RNA molecule (Schuster-Gossler *et al* 1998, Developmental Dynamics vol 212 pp214-28). It has been shown recently that such RNA molecules may control directly the expression of imprinted genes (Sleutels 2002, Nature vol 415 pp810-813). Our discovery that this gene is inducible by the hypoxia-related transcription factor EPAS1 raises the likelihood of its involvement in cell survival under stress.

The protein sequence encoded by hypothetical protein KIAA1696 is represented in the public databases by the accession XP_051010 and is described herein as SEQ ID No. 3. The nucleotide sequence is represented in the public sequence databases by the accession XM_051010 and is described herein as SEQ ID No. 4. KIAA1696 contains a PHD Zinc finger motif that is commonly found in nuclear proteins thought to be involved in chromatin-mediated transcriptional regulation. Thus our observation of EPAS1--inducibility enables the definition of a role for this protein in the regulation of genes necessary for cell survival under conditions of oxygen deprivation.

The protein sequence encoded by EGLN3 is represented in the public databases by the accession NP_071356 and is described herein as SEQ ID No. 5. The nucleotide sequence is represented in the public sequence databases by the accession NM_022073 and is described herein as SEQ ID No. 6. EGLN3 is a member of the hif prolyl hydroxylase family of enzymes, which are involved in the degradation of transcription factors such as Hif1 alpha (Bruick *et al* 2001, Science vol 294 pp1337-40). The prolyl hydroxylation activity of EGLN3 can be inhibited by the use of oxogluterate analogs. The residues His 135 and Asp 137 are key residues in the active site of the enzyme and therefore mutation of either of these residues will lead to a loss of function. Our observation of the strong inducibility of this gene by EPAS1 enables the definition of its role in the hypoxic response, providing the means for rapid adaptation of cells to normoxic conditions.

The protein sequence encoded by the splice variant of PHPS1 is represented in the public databases by the accession XP_096898 and is described herein as the SEQ ID No. 7. The nucleotide sequence is represented in the public sequence databases by the accession BE671816 and is described herein as SEQ ID No. 8. The protein encoded by the splice variant of PHPS1 contains a transmembrane domain and has an unknown function. Our observation of its Hif1α-inducibility enables the definition of its role in cell survival under conditions of oxygen deprivation. Its transmembrane domain increases the likelihood of effective therapeutic intervention.

SEQ ID No. 10 corresponds to a singleton EST that is represented in the public databases by the accession N68173. Our observation of its EPAS1-inducibility enables the definition of its role in cell survival under conditions of oxygen deprivation.

The protein sequence encoded by IL6R is represented in the public databases by the accession NP_000556 and is described herein as SEQ ID No. 11. The nucleotide sequence is represented in the public sequence databases by the accession NM_000565 and is described herein as SEQ ID No. 12. Elevated levels of IL6 and the IL6 receptor (IL6R) have recently been shown to be associated with cardiac disease (Plenz *et al* 2001, Eur J Heart Failure vol 3 pp415-21). Also, the ablation of IL6R activity is known to be effective in the treatment of inflammatory conditions (Nishimoto *et al* 2000, Annals of the Rheumatic Diseases 2000 vol 59 Suppl 1 pp21-7). Our observation of the EPAS1-inducibility of IL6R enables the prediction of its hypoxia-inducibility. The artificial control of the hypoxic response could reduce the level of IL6R in relevant tissues and thus ameliorate the effects of IL6R on disease processes.

The protein sequence encoded by hypothetical protein XP_091988 is represented in the public databases by the accession XP_091988 and is described herein as SEQ ID No. 13. The nucleotide sequence is represented in the public sequence databases by the accession XM_091988 and is described herein as SEQ ID No. 14. Hypothetical protein XP_091988 represents a Zinc finger protein likely to be involved in protein-protein interactions via its KRAB domain. Thus our observation of EPAS1-inducibility enables the definition of a role for this protein in the regulation of genes necessary for cell survival under conditions of oxygen deprivation.

SEQ ID No. 18 represents a singleton EST that is present in the public sequence databases with the accession number R95132. Our observation of its EPAS1-inducibility enables the definition of its role in cell survival under conditions of oxygen deprivation.

The protein sequence encoded by FRA2, also known as Fos-like antigen 2, is represented in the public databases by the accession P15408 and is described herein as SEQ ID No. 19. The nucleotide sequence is represented in the public sequence databases by the accession BC022791 and is described herein as SEQ ID No. 20. FRA2 is a DNA binding Leucine zipper and is a member of the FOS superfamily. Members of the FOS superfamily have been implicated as regulators of cell proliferation, differentiation and transformation and are known to dimerize with members of the Jun family of Leucine zippers to form the AP-1 complex. FRA2, along with other AP-1 proteins, shows a correlation of elevated levels with clinico-pathologic tumour parameters. It is implicated in the regulation of cell cycle progression (Bamberger *et al* 2001, J Cancer Res Clin Oncol vol 127 pp545-50) and has been shown to be up-regulated in immortalized cells (Sheerin *et al.* 2002 Cancer Lett Mar 8; 177(1):83-7). The effects of FRA2 on differentiation are well known; e.g. c-Fos and Fra2 co-operate with c-Jun to prevent terminal myoblast differentiation and withdrawal from the cell cycle (Daury *et al* 2001, Oncogene vol 20 pp 7998-8008). FRA2 is also required for cytokine gene induction and is involved in the regulation of TNF production (Udalova *et al.* 2001 Biochem Biophys Res Commun, vol 289 pp25-33). Levels of FRA2 have been shown to decrease with age (Medicherla *et al.* 2001, Mech Ageing Dev, vol 122 pp169-86). FRA2 is highly up-regulated in epithelial cells under hypoxic stress. FRA2 also shows marked elevation in repeatedly stressed rates. FRA2 up-regulates the stress-related enzymes tyrosine hydroxylase and dopamine beta-hydroxylase (Nakova *et al* 2000, J Neuroscience vol 20 pp5647-53). FRA2 has also been implicated with a role in tissue remodelling. TIMP is up-regulated in response to cell damage. FRA2 plays a role in that up-regulation in injured liver hepatic stellate cells (Bahr *et al* 1999, Hepatology vol 29 pp839-48). SPARC is an extracellular matrix protein, which is thought to contribute to the remodelling of the extracellular matrix during neoplastic progression. Jun-mediated repression of SPARC is enhanced by FRA2 (Vial *et al* 2000, Oncogene vol 19 pp5020-9). Our observation of its EPAS1-inducibility enables the definition of its role under conditions of oxygen deprivation, in the regulation of genes necessary for cell survival. These conditions apply within solid tumours, and in tumour cells cell cycle progression is a crucial pathologic feature. Tissue re-modelling is potentially a damaging process in diseases such as rheumatoid arthritis. Thus we have shown that interfering with the hypoxia-induction of FRA2 could have a beneficial therapeutic effect.

The full-length protein sequence encoded by DBI, also known as Membrane-associated diazepam binding inhibitor, is not represented in the public database. A fragment of the protein is represented by accession number XP_038526 and is described here by the SEQ ID No. 21.

The full-length nucleotide sequence of DBI is also absent from the public databases but is described here by SEQ ID No. 22. DBI binds to an allosteric regulatory site and reduces the activity of the GABA(A) receptor complex. (Gray *et al* 1986, PNAS vol 83 pp7547-51. Bormann 1991, Neuropharmacology vol 30 pp1387-9.) GABA(A) activity regulates steroidogenesis, and thus has a role in reproductive hormone secretion (Papadopoulos *et al* 1991, Neuropharmacology vol 30 pp1417-23), and also in the regulation of the behavioural responses to stress via corticotrophin releasing factor (CRF) (Guidotti 1991, Neuropharmacology vol 30 npp1425-33). Very recently it has been shown that CRF has utility as an anti-inflammatory agent (McLoon *et al* 2002, Inflamm Res vol 51 pp16-23). Our observation of the Hif1α and EPAS1 regulation of DBI indicates the utility of intervention via DBI to influence hypoxia-induced apsects of the stress response and inflammation. DBI binds acyl-CoA esters, and affinity increases with acyl chain length (Knusden 1991, Neuropharmacology vol 30 pp1405-10). In October 2001 a paper was published linking co-incident expression of DBI and the fatty acid binding protein (FAPB). We have shown previously that expression of FAPB is induced by hypoxia (see UK application 0030076.4). The tissue distribution of DBI encompasses regions outside the brain, specifically squamous epithelia in the digestive system (Yanase *et al* 2001, Arch Histol Cytol vol 64 pp449-60). The coordinate hypoxia induction of DBI and FAPB implies a joint role in cell survival under conditions of oxygen deprivation.

The protein sequence encoding BACH2 is represented in the public databases by the accession CAC28130 and is described herein as SEQ ID No. 23. The nucleotide sequence is represented in the public databases by the accession AJ271878 and is described herein as SEQ ID No. 24. Bach2 is a transcriptional regulator containing a basic leucine zipper domain. Its synthesis is transcriptionally repressed by constitutive expression of the oncogene BCR/ABL, a tyrosine kinase. It has been shown recently that BACH2 functions as a tumour suppressor by repressing transcription of genes that lead to progression of some leukaemias (Vieira *et al* 2001, Genes Chromosomes Cancer vol 32 pp353-63). The antiproliferative effect of BACH2 could be harnessed in the treatment of malignant disease. Our observation that this gene is induced by EPAS1 makes clear that artificial control of the hypoxic response could have an antiproliferative effect in some cell types as a result of increased BACH2 activity.

SEQ ID No. 26 represents a singleton EST that is present in the public sequence databases with the accession number T8601. Our observation of its EPAS1-inducibility enables the definition of its role in cell survival under conditions of oxygen deprivation.

The protein sequence encoding Aldehyde dehydrogenase 12 is represented in the public sequence database by the accession NP_072090 and is described herein as SEQ ID No. 27. The nucleotide sequence is represented in the public databases by the accession NM_022568 and is described herein as SEQ ID No. 28. Our observation of its EPAS1-inducibility enables the definition of its role in cell survival under conditions of oxygen deprivation.

SEQ ID No. 30 represents a singleton EST that is present in the public sequence databases with the accession number R06745. Our observation of its Hif1α and EPAS1-inducibility enables the definition of its role in cell survival under conditions of oxygen deprivation.

SEQ ID No. 32 represents a singleton EST that is present in the public sequence databases with the accession number N64734. Our observation of its strong EPAS1-inducibility enables the definition of its role in cell survival under conditions of oxygen deprivation.

SEQ ID No. 34 represents a singleton EST that is present in the public sequence databases with the accession number T85201. Our observation of its strong EPAS1-inducibility enables the definition of its role in cell survival under conditions of oxygen deprivation.

The protein sequence encoded by the rat metallothionein 1a gene is represented in the public databases by the accession AAA41590 and is described herein as SEQ ID No. 37. The nucleotide sequence is represented in the public sequence databases by the accession J00750 and is described herein as SEQ ID No. 38. The metallothionein genes are known to play a role in scavenging free radicals and therefore protecting against DNA damage (Irato *et al* 2001, Immunol Cell Biol vol 79 pp251-4). Our observation of its Hif1α -inducibility enables the definition of its role in cell survival under conditions of oxygen deprivation.

References

Blackstock and Weir (1999) *Trends in Biotech.* 17: 121-126.

Bowtell (1999) *Nature Genetics* 21: 25-32.

Eisen and Brown (1999). *Methods Enymol.* 303: 179-205.

Griffiths L. *et al.* (2000), *Gene Therapy* 7: 255-262.

Jorgensen *et al.* (1999), *Electrophoresis* 2: 230-40.

Kirschbaum and Kozian (1999), *Trends in Biotech* 17: 73-78.

Mitophanous *et al.* (1999), *Gene Therapy* 6: 1808-1818

Pardee and Liang (1992), *Science* 257: 967-971.

Rabilloud *et al.* (1997), *Electrophoresis* 18: 307-316.

Soneoka *et al.* (1995, *Nucleic Acids Res.* 23: 628-33

Wilkinson *et al.* (1995), *Plant Mol Biol* 6: 1097-108.

Zhang *et al.* (1998), *Mol Biotechnol* 10(2): 155-65.

Zhao *et al.* (1999), *J Biotechnol.* 73(1): 35-41.

CLAIMS

1. A substantially purified polypeptide, which polypeptide:

   i) comprises or consists of the amino acid sequence as recited in any one of SEQ ID Nos: 3, 5, 7, 11, 13, 19, 21, 23 or 27;

   ii) comprises or consists of an amino acid sequence encoded by a nucleic acid sequence recited in any one of SEQ ID Nos: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 or 36, or encoded by a gene identified from an EST recited in any one of these SEQ ID Nos;

   iii) is a fragment of a polypeptide according to i) or ii), provided that said fragment retains a biological activity possessed by the full length polypeptide of i) or ii), or has an antigenic determinant in common with the polypeptide of i) or ii); or

   iv) is a functional equivalent of a polypeptide of i), ii) or (iii).

2. A polypeptide according to claim 1, wherein said biological activity is a hypoxia-regulated activity.

3. A polypeptide according to claim 2, wherein the expression of the polypeptide is hypoxia-induced.

4. A polypeptide which is a functional equivalent according to part iv) of any one of claims 1-3, is homologous to the amino acid sequence as recited in any one of SEQ ID Nos: 3, 5, 7, 11, 13, 19, 21, 23 or 27 or is homologous to the amino acid sequence encoded by a nucleic acid as recited in any one of SEQ ID Nos: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 or 36, and has equivalent biological activity to that possessed by the full length polypeptide of i) or ii).

5. A fragment or functional equivalent according to any one of claims 1-4, which has greater than 50% sequence identity with the amino acid sequence as recited in any one of SEQ ID Nos: 3, 5, 7, 11, 13, 19, 21, 23 or 27 or with the amino acid sequence that is encoded by a nucleic acid as recited in any one of SEQ ID Nos: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 or 36, or with fragments thereof, preferably greater than 60%, 70%, 80%, 90%, 95%, 98% or 99% sequence identity.

6. A fragment as recited in any one of claims 1-5, having an antigenic determinant in common with a polypeptide according to part i) of any one of claims 1-5, which consists of

7 or more (for example, 8, 10, 12, 14, 16, 18, 20 or more) amino acid residues from the amino acid sequence as recited in any one of SEQ ID Nos: 3, 5, 7, 11, 13, 19, 21, 23 or 27 or the amino acid sequence encoded by a nucleic acid as recited in any one of SEQ ID Nos: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 or 36.

7. A purified and isolated nucleic acid molecule that encodes a polypeptide according to any one of claims 1-6.

8. A purified nucleic acid molecule according to claim 7, which consists of the nucleic acid sequence as recited in any one of SEQ ID Nos.: 2, 4, 6, 8, 10, 12, 14, 18, 20, 22, 24, 26, 28, 30, 32, 34 or 36, or is a redundant equivalent or fragment thereof.

9. A purified nucleic acid molecule which hydridizes under high stringency conditions with a nucleic acid molecule according to claim 7 or claim 8.

10. A vector comprising a nucleic acid molecule as recited in any one of claims 7-9.

11. A delivery vehicle comprising a nucleic acid according to any one of claims 7-9, or a vector according to claim 10.

12. A host cell transformed with a vector according to claim 11.

13. An antagonist ligand which binds specifically to, and which inhibits the hypoxia-induced activity of, a polypeptide according to any one of claims 1-6.

14. An agonist ligand which binds specifically to, and which activates the hypoxia-induced activity of, a polypeptide according to any one of claims 1-6 to augment or potentiate a hypoxia-induced activity.

15. A ligand according to claim 13 or 14, which is an antibody.

16. A ligand according to claim 13 or 14, which is a peptide, a peptidomimetic, or a drug molecule, such as a small natural or synthetic organic molecule of up to 2000Da, preferably 800Da or less.

17. A pharmaceutical composition suitable for modulating hypoxia and/or ischaemia, comprising a therapeutically-effective amount of a polypeptide according to any one of claims 1-6, a nucleic acid molecule according to any one of claims 7-9, a vector according to claim 10 or a ligand according to any one of claims 13-16, in conjunction with a pharmaceutically-acceptable carrier.

18. A vaccine composition comprising a polypeptide according to any one of claims 1-6, a nucleic acid molecule according to any one of claims 7-9, a vector according to claim 10 or a ligand according to any one of claims 13-16.

19. A polypeptide according to any one of claims 1-6, a nucleic acid molecule according to any one of claims 7-9, a vector according to claim 10 or a ligand according to any one of claims 13-16, for use in therapy or diagnosis of disease.

20. A polypeptide, nucleic acid molecule, vector or ligand as recited in claim 19, wherein said disease is a hypoxia-regulated condition.

21. A polypeptide, nucleic acid molecule, vector or ligand as recited in claim 19, wherein said hypoxia-regulated condition is tumourigenesis, angiogenesis, apoptosis, inflammation, erythropoiesis, the biological response to hypoxia conditions (including processes such as glycolysis, gluconeogenesis, glucose transportation, catecholamine synthesis, iron transport or nitric oxide synthesis).

22. A method of treating a disease in a patient in need of such treatment by administering to a patient a therapeutically effective amount of a polypeptide according to any one of claims 1-6, a nucleic acid molecule according to any one of claims 7-9, a vector according to claim 10, a ligand according to any one of claims 13-16 or a pharmaceutical composition according to claim 17.

23. A method of regulating tumourigenesis, angiogenesis, apoptosis, the biological response to hypoxia conditions, or a hypoxic-associated pathology in a patient in need of such treatment by administering to a patient a therapeutically effective amount of a polypeptide according to any one of claims 1-6, a nucleic acid molecule according to any one of claims 7-9, a vector according to claim 10, a ligand according to any one of claims 13-16 or a pharmaceutical composition according to claim 17.

24. A method according to claim 23, wherein, for diseases in which the expression of the natural gene or the activity of the polypeptide is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand, compound or composition administered to the patient is an agonist.

25. A method according to claim 23, wherein, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an antagonist.

26. A polypeptide according to any one of claims 1-6, a nucleic acid molecule according to any one of claims 7-9, a vector according to claim 10, a ligand according to any one of claims 13-16 or a pharmaceutical composition according to claim 17, for use in the manufacture of a medicament for the treatment of a hypoxia-regulated condition.

27. A method of monitoring the therapeutic treatment of a disease or physiological condition in a patient, comprising monitoring over a period of time the level of expression or activity of a polypeptide according to any one of claims 1-6, a nucleic acid molecule according to any one of claims 7-9, a vector according to claim 10, a ligand according to any one of claims 13-16 or a pharmaceutical composition according to claim 17, in tissue from said patient, wherein altering said level of expression or activity over the period of time towards a control level is indicative of regression of said disease.

28. A method of providing a hypoxia regulating gene, an apoptotic or an angiogenesis regulating gene by administering directly to a patient in need of such therapy an expressible vector comprising expression control sequences operably linked to one or more of the nucleic acid molecules recited in claims 7-9.

29. A method of diagnosing a hypoxia-regulated condition in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide according to any one of claims 1-6, or assessing the activity of such a polypeptide, in tissue from said patient and comparing said level of expression or activity to a control level, wherein a level that is different to said control level is indicative of the hypoxia-related condition.

30. A method according to claim 29, that is carried out *in vitro*.

31. A method according to claim 29 or claim 30, which comprises the steps of: (a) contacting a ligand according to any one of claims 13-16 with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

32. A method according to claim 29 or claim 30, comprising the steps of:

a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule according to any one of claims 7-9 and the probe;

b) contacting a control sample with said probe under the same conditions used in step a); and

c) detecting the presence of hybrid complexes in said samples;

wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid complex in the control sample is indicative of the hypoxia-related condition.

33. A method according to claim 29 or claim 30, comprising the steps of:

a) contacting a sample of nucleic acid from tissue of the patient with a nucleic acid primer under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule according to any one of claims 7-9 and the primer;

b) contacting a control sample with said primer under the same conditions used in step a);

c) amplifying the sampled nucleic acid; and

d) detecting the level of amplified nucleic acid from both patient and control samples;

wherein detection of levels of the amplified nucleic acid in the patient sample that differ significantly from levels of the amplified nucleic acid in the control sample is indicative of the hypoxia-related condition.

34. A method according to claim 29 or claim 30, comprising the steps of:

a) obtaining a tissue sample from a patient being tested for the hypoxia-related condition;

b) isolating a nucleic acid molecule according to any one of claims 7-9 from said tissue sample; and

c) diagnosing the patient for disease by detecting the presence of a mutation which is associated with the hypoxia-related condition in the nucleic acid molecule as an indication of the hypoxia-related condition.

35. The method of claim 34, further comprising amplifying the nucleic acid molecule to form an amplified product and detecting the presence or absence of a mutation in the amplified product.

36. A method according to any one of claims 29-35, wherein said disease is cancer, ischaemic conditions, reperfusion injury, retinopathy, neonatal stress, preeclapmsia, atherosclerosis, inflammatory conditions, wound healing, tumourigenesis, angiogenesis, apoptosis, inflammation or erythropoiesis.

37. A method according to claim 36, wherein said hypoxia or ischaemia-related tissue damage is due to a disorder of the cerebral, coronary or peripheral circulation.

38. A method according to any one of claims 29, 30 and 32-36, wherein the tissue is a cancer tissue.

39. A method for the identification of a compound that is effective in the treatment and/or diagnosis of disease, comprising contacting a polypeptide according to any one of claims 1-6, or a nucleic acid molecule according to any one of claims 7-9 with one or more compounds suspected of possessing binding affinity for said polypeptide or nucleic acid molecule, and selecting a compound that binds specifically to said nucleic acid molecule or polypeptide.

40. A compound identified by a method according to claim 39.

41. A compound according to claim 40, which is a natural or modified substrate, an enzyme, a receptor, a small organic molecule, such as a small natural or synthetic organic molecule of up to 2000Da, preferably 800Da or less, a peptidomimetic, an inorganic molecule, a peptide, a polypeptide, an antibody, or a structural or functional mimetics of any of these compounds.

42. A kit useful for diagnosing disease comprising a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to any one of claims 7-9; a second container containing primers useful for amplifying said nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease.

43. The kit of claim 42, further comprising a third container holding an agent for digesting unhybridised RNA.

44. An array of at least two nucleic acid molecules, wherein each of said nucleic acid molecules either corresponds to the sequence of, is complementary to the sequence of, or hybridises specifically to a nucleic acid molecule according to any one of claims 7-9.

45. An array according to claim 44, which contains nucleic acid molecules that either correspond to the sequence of, are complementary to the sequence of, or hybridise specifically to at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 or 18 of the nucleic acid molecules implicated in a hypoxia-regulated condition as recited in claims 7-9.

46. An array according to any claim 44 or claim 45, wherein said nucleic acid molecules consist of between twelve and two thousand nucleotides.

47. An array of antibodies, comprising at least two different antibody species, wherein each antibody species is immunospecific with a polypeptide implicated in a hypoxia-regulated condition as recited in any one of claims 1-6.

48. An array of polypeptides, comprising at least two polypeptide species as recited in any one of claims 1-6, wherein each polypeptide species is implicated in a hypoxia-regulated condition, or is a functional equivalent variant or fragment thereof.

49. A kit comprising an array of nucleic acid molecules according to any one of claims 7-9.

50. A kit comprising one or more antibodies that bind to a polypeptide as recited in any one of claims 1-6; and a reagent useful for the detection of a binding reaction between said antibody and said polypeptide.

51. A transgenic or knockout non-human animal that has been transformed to express higher, lower or absent levels of a polypeptide according to any one of claims 1-6.

52. A method for screening for a compound effective to treat disease, by contacting a non-human transgenic animal according to claim 51 with a candidate compound and determining the effect of the compound on the disease of the animal.

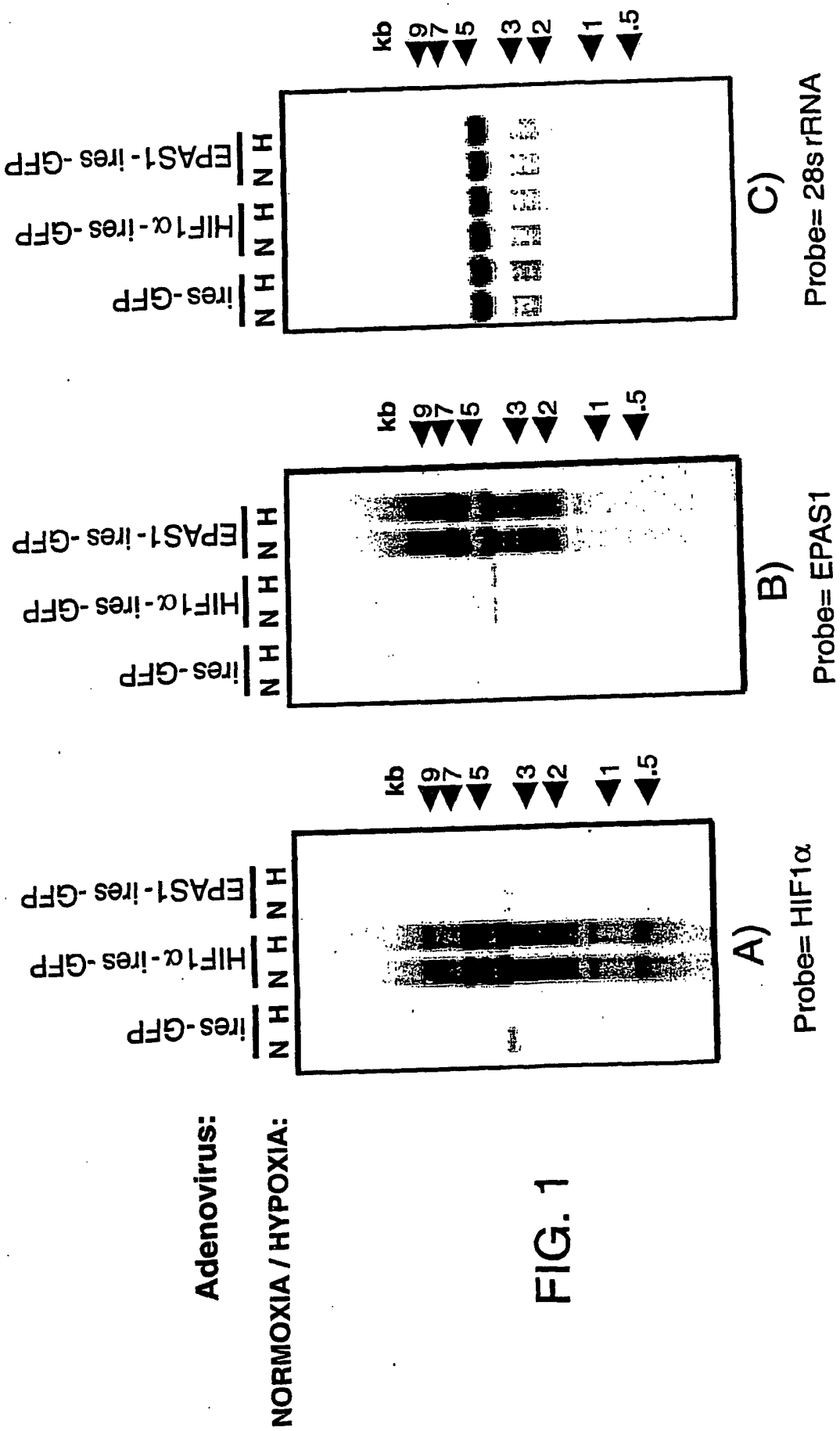

Adenovirus:
NORMOXIA / HYPOXIA:
N H N H N H
ires-GFP
HIF1α-ires-GFP
EPAS1-ires-GFP
kb
9 7 5 3 2 1 .5
A)
Probe= HIF1α
B)
Probe= EPAS1
C)
Probe= 28s rRNA

FIG. 1

FIG. 2

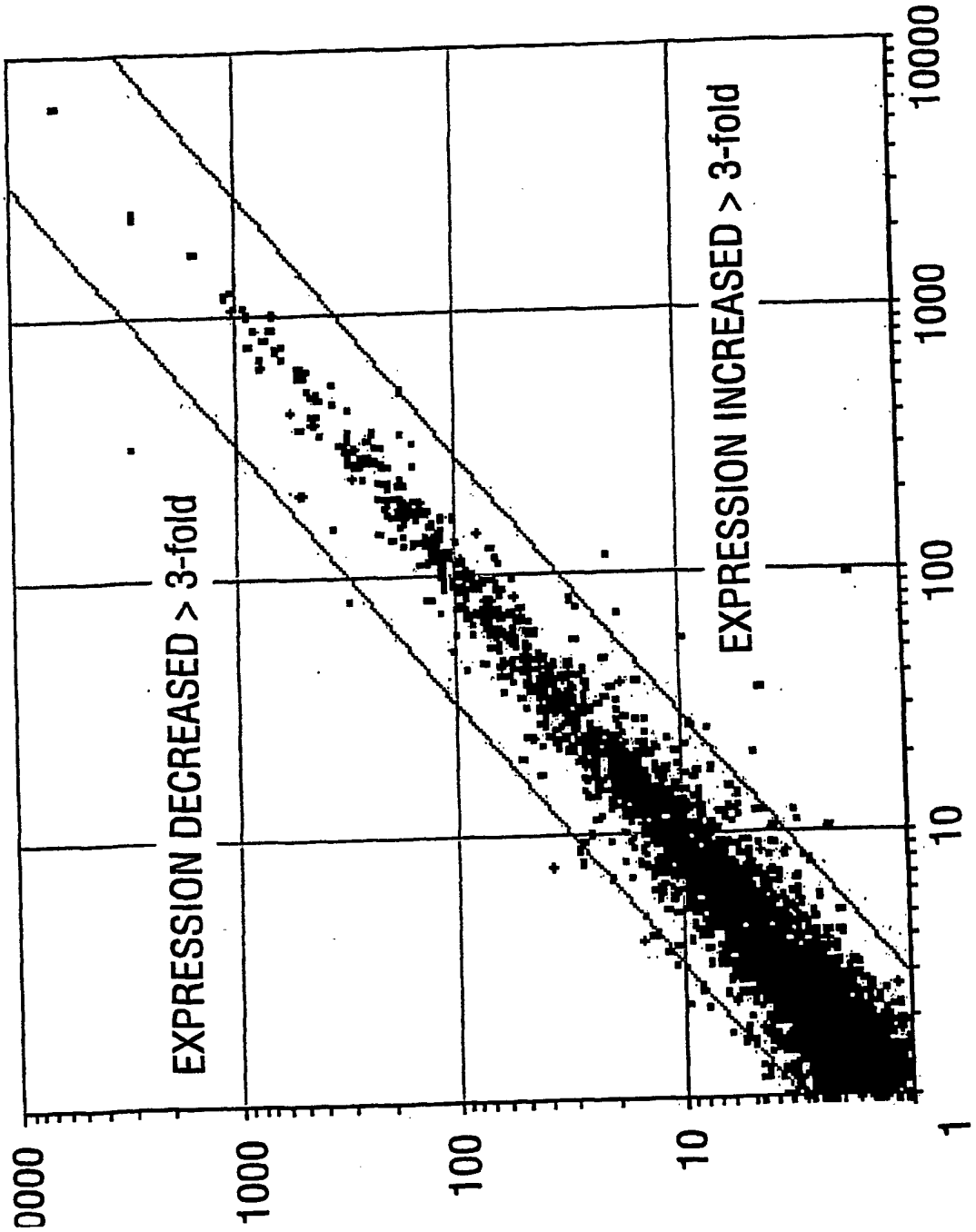

EXPRESSION DECREASED > 3-fold
EXPRESSION INCREASED > 3-fold
10000
1000
100
10
1
0000
1000
100
10

FIG. 3A

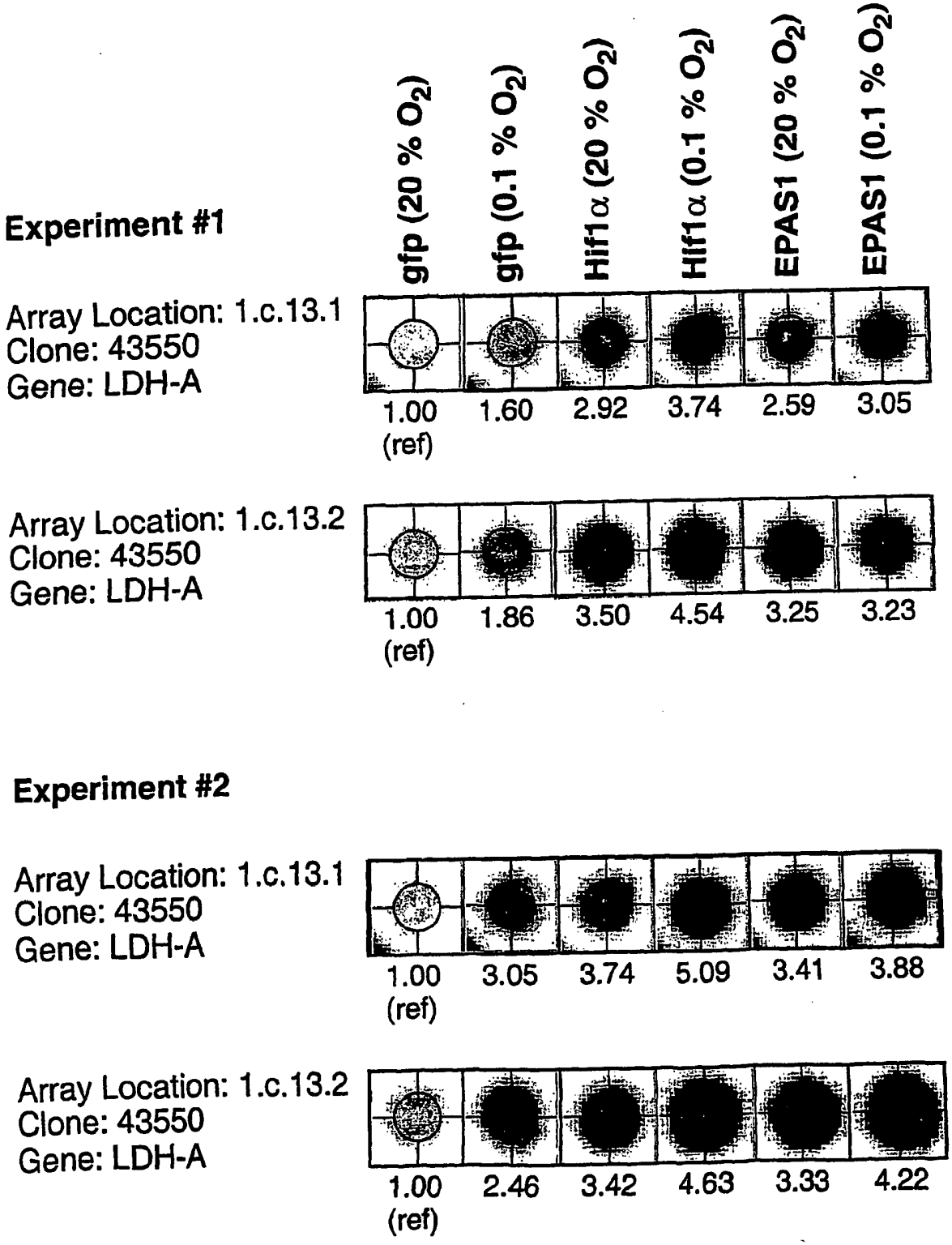

gfp (20 % O2)
gfp (0.1 % O2)
Hif1α (20 % O2)
Hif1α (0.1 % O2)
EPAS1 (20 % O2)
EPAS1 (0.1 % O2)
Experiment #1
Array Location: 1.c.13.1
Clone: 43550
Gene: LDH-A
1.00 (ref)
1.60
2.92
3.74
2.59
3.05
Array Location: 1.c.13.2
Clone: 43550
Gene: LDH-A
1.00 (ref)
1.86
3.50
4.54
3.25
3.23
Experiment #2
Array Location: 1.c.13.1
Clone: 43550
Gene: LDH-A
1.00 (ref)
3.05
3.74
5.09
3.41
3.88
Array Location: 1.c.13.2
Clone: 43550
Gene: LDH-A
1.00 (ref)
2.46
3.42
4.63
3.33
4.22

FIG. 3B

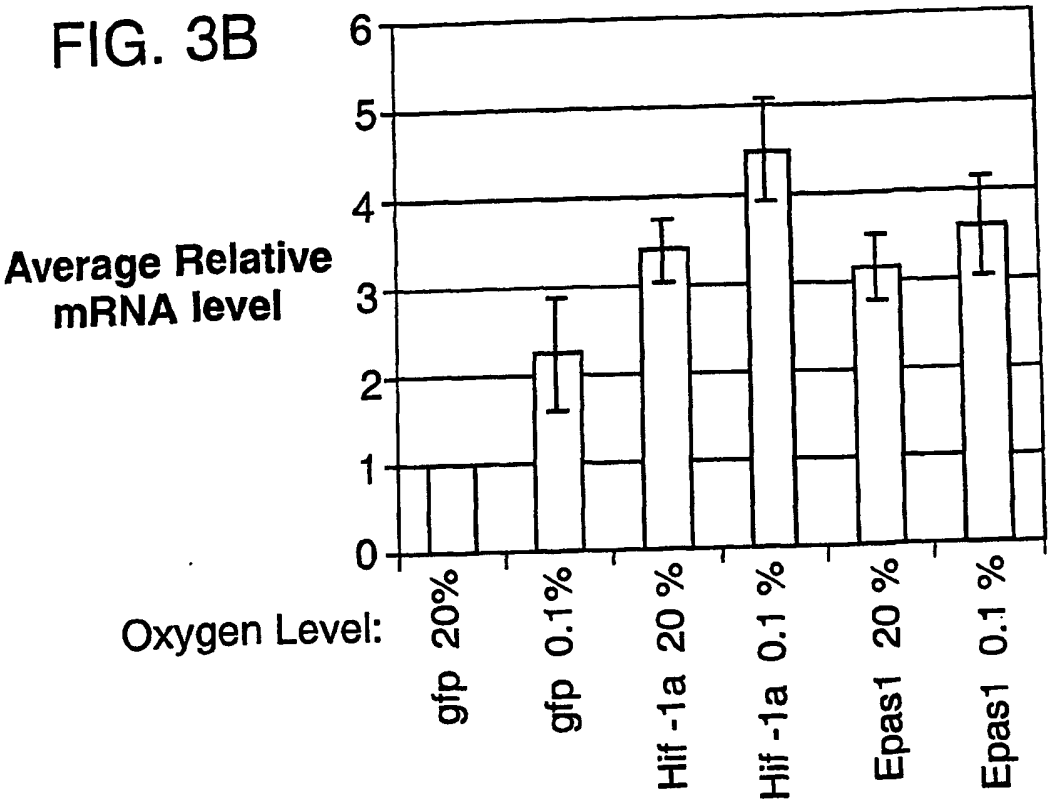

Average Relative mRNA level
0
1
2
3
4
5
6
Oxygen Level:
gfp 20%
gfp 0.1%
Hif -1a 20 %
Hif -1a 0.1 %
Epas1 20 %
Epas1 0.1 %

FIG. 4B

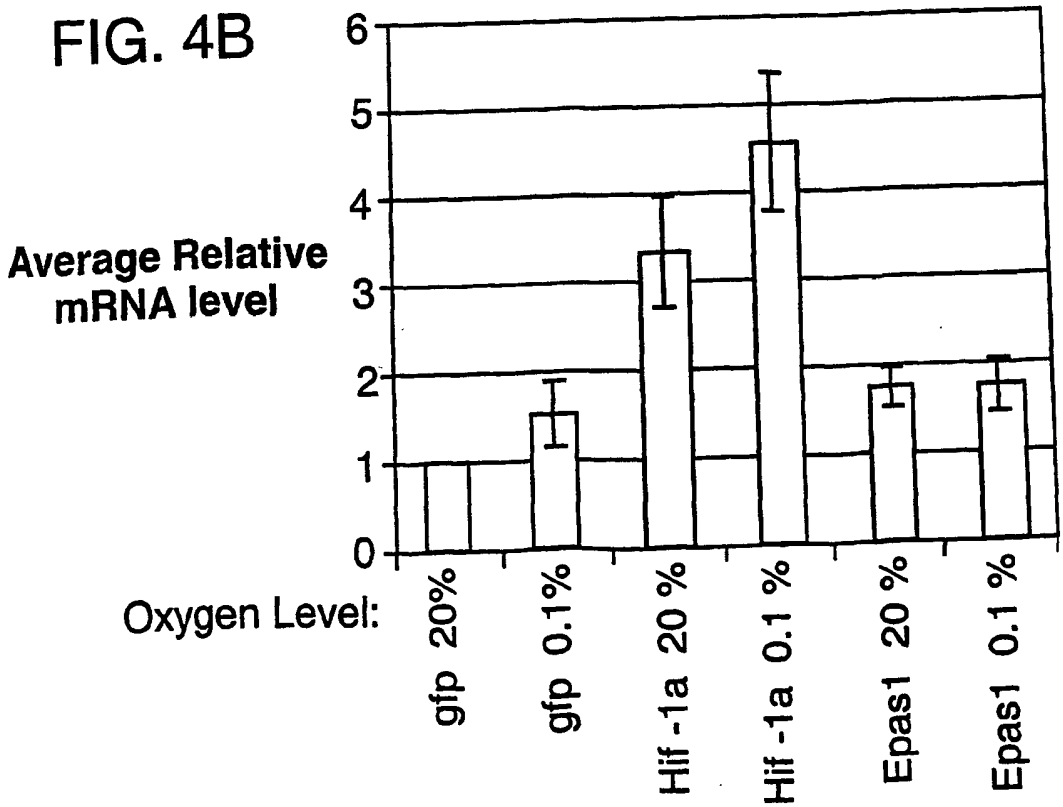

Average Relative mRNA level
0
1
2
3
4
5
6
Oxygen Level:
gfp 20%
gfp 0.1%
Hif -1a 20 %
Hif -1a 0.1 %
Epas1 20 %
Epas1 0.1 %

FIG. 4A

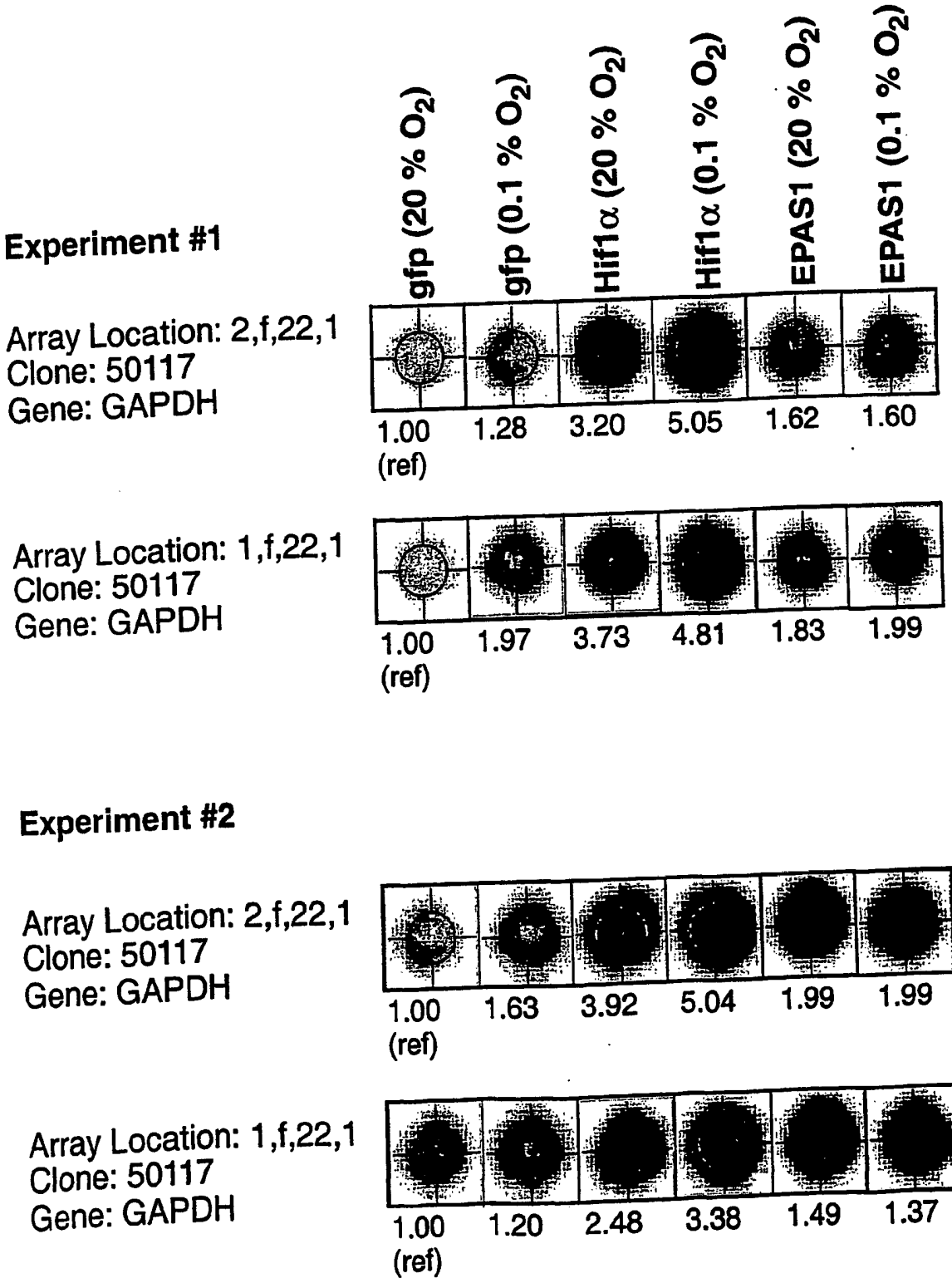

gfp (20 % O2)
gfp (0.1 % O2)
Hif1α (20 % O2)
Hif1α (0.1 % O2)
EPAS1 (20 % O2)
EPAS1 (0.1 % O2)
Experiment #1
Array Location: 2,f,22,1
Clone: 50117
Gene: GAPDH
1.00 (ref)
1.28
3.20
5.05
1.62
1.60
Array Location: 1,f,22,1
Clone: 50117
Gene: GAPDH
1.00 (ref)
1.97
3.73
4.81
1.83
1.99
Experiment #2
Array Location: 2,f,22,1
Clone: 50117
Gene: GAPDH
1.00 (ref)
1.63
3.92
5.04
1.99
1.99
Array Location: 1,f,22,1
Clone: 50117
Gene: GAPDH
1.00 (ref)
1.20
2.48
3.38
1.49
1.37

FIG. 5A

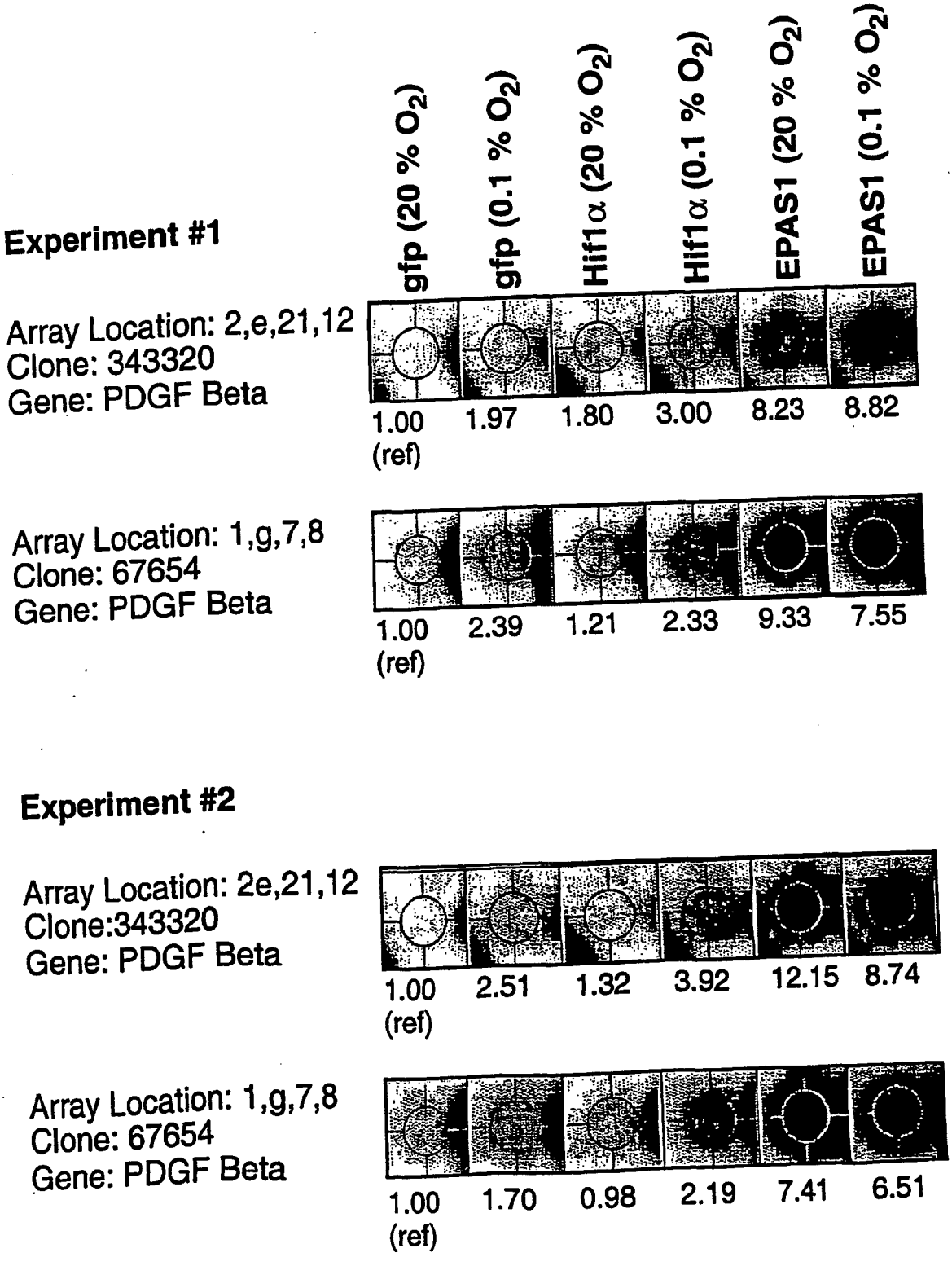

gfp (20 % O2)
gfp (0.1 % O2)
Hif1α (20 % O2)
Hif1α (0.1 % O2)
EPAS1 (20 % O2)
EPAS1 (0.1 % O2)
Experiment #1
Array Location: 2,e,21,12
Clone: 343320
Gene: PDGF Beta
1.00 (ref)
1.97
1.80
3.00
8.23
8.82
Array Location: 1,g,7,8
Clone: 67654
Gene: PDGF Beta
1.00 (ref)
2.39
1.21
2.33
9.33
7.55
Experiment #2
Array Location: 2e,21,12
Clone:343320
Gene: PDGF Beta
1.00 (ref)
2.51
1.32
3.92
12.15
8.74
Array Location: 1,g,7,8
Clone: 67654
Gene: PDGF Beta
1.00 (ref)
1.70
0.98
2.19
7.41
6.51

FIG. 5B

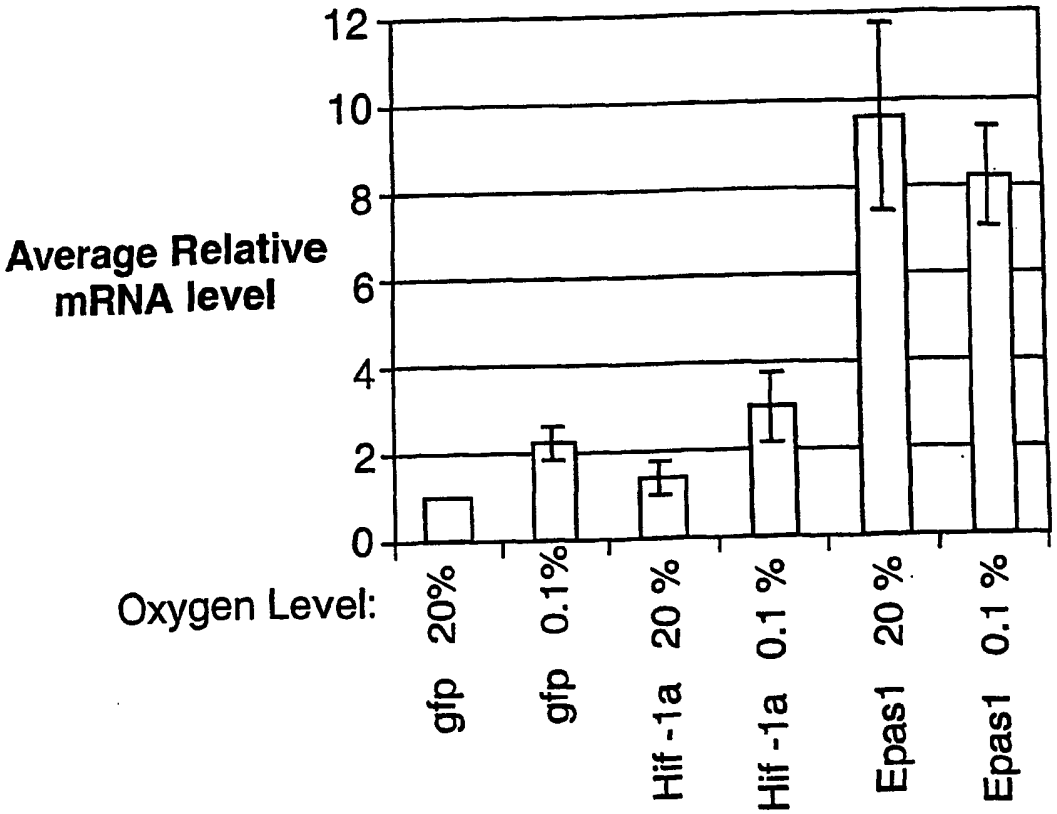

Average Relative mRNA level
12
10
8
6
4
2
0
Oxygen Level:
gfp 20%
gfp 0.1%
Hif -1a 20 %
Hif -1a 0.1 %
Epas1 20 %
Epas1 0.1 %

FIG. 6A

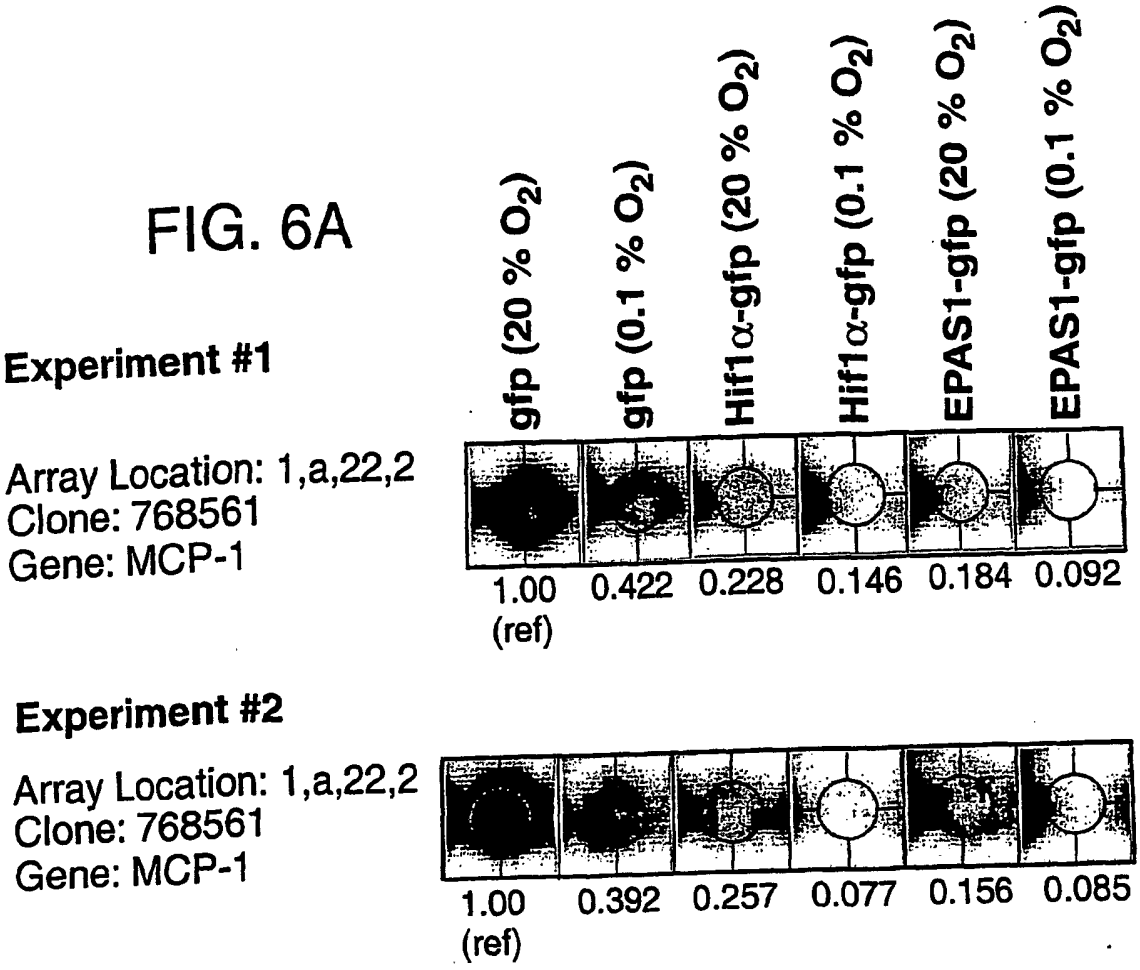

Experiment #1
Array Location: 1,a,22,2
Clone: 768561
Gene: MCP-1
gfp (20 % O2)
gfp (0.1 % O2)
Hif1α-gfp (20 % O2)
Hif1α-gfp (0.1 % O2)
EPAS1-gfp (20 % O2)
EPAS1-gfp (0.1 % O2)
1.00 (ref)
0.422
0.228
0.146
0.184
0.092
Experiment #2
Array Location: 1,a,22,2
Clone: 768561
Gene: MCP-1
1.00 (ref)
0.392
0.257
0.077
0.156
0.085

FIG. 6B

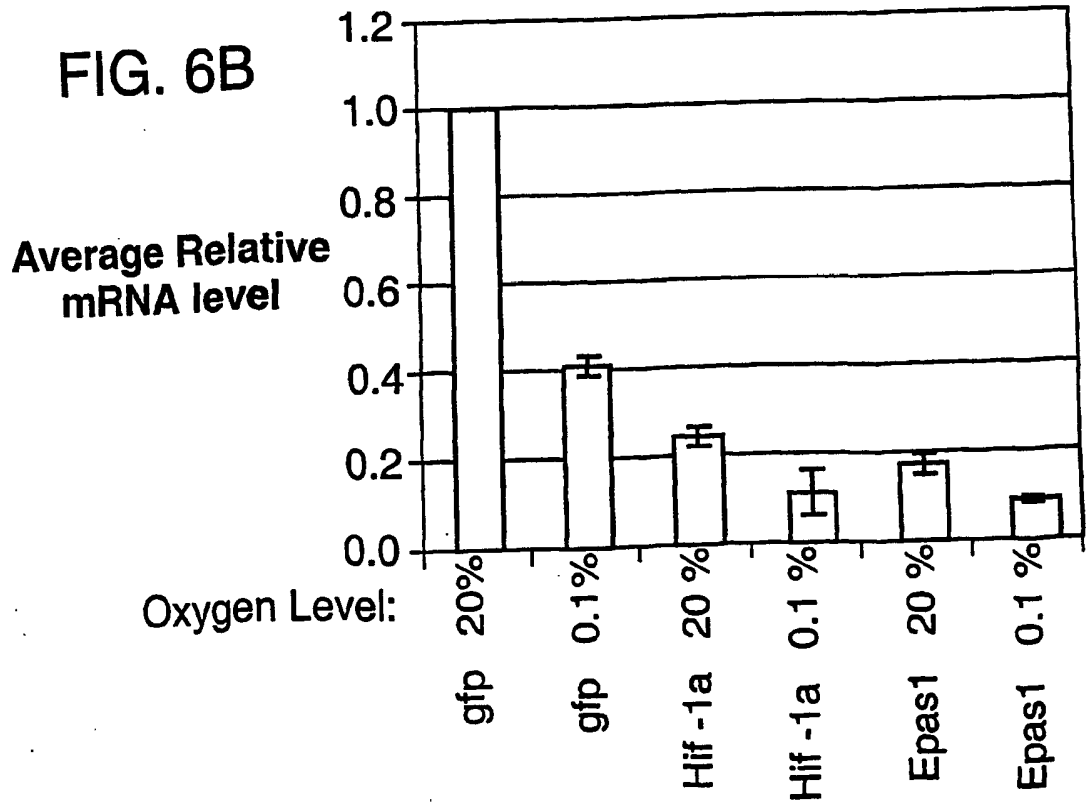
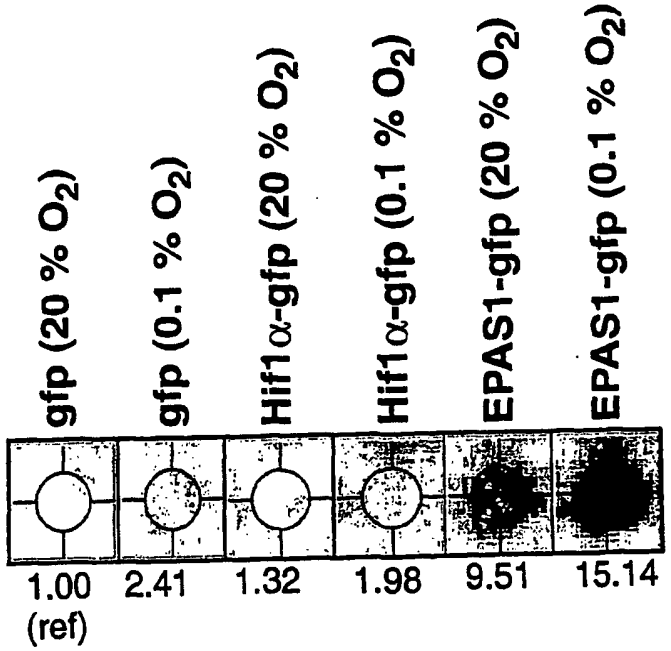

Average Relative mRNA level
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Oxygen Level:
20%
0.1%
20 %
0.1 %
20 %
0.1 %
gfp
gfp
Hif -1a
Hif -1a
Epas1
Epas1

FIG. 7A

Experiment #1

Array Location: 2,a,30,5
Clone: 293336
Gene: (only ESTs)



Experiment #2

Array Location: 2,a,30,5
Clone: 293336
Gene: (only ESTs)

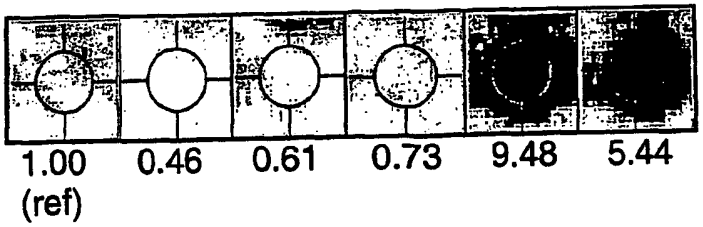
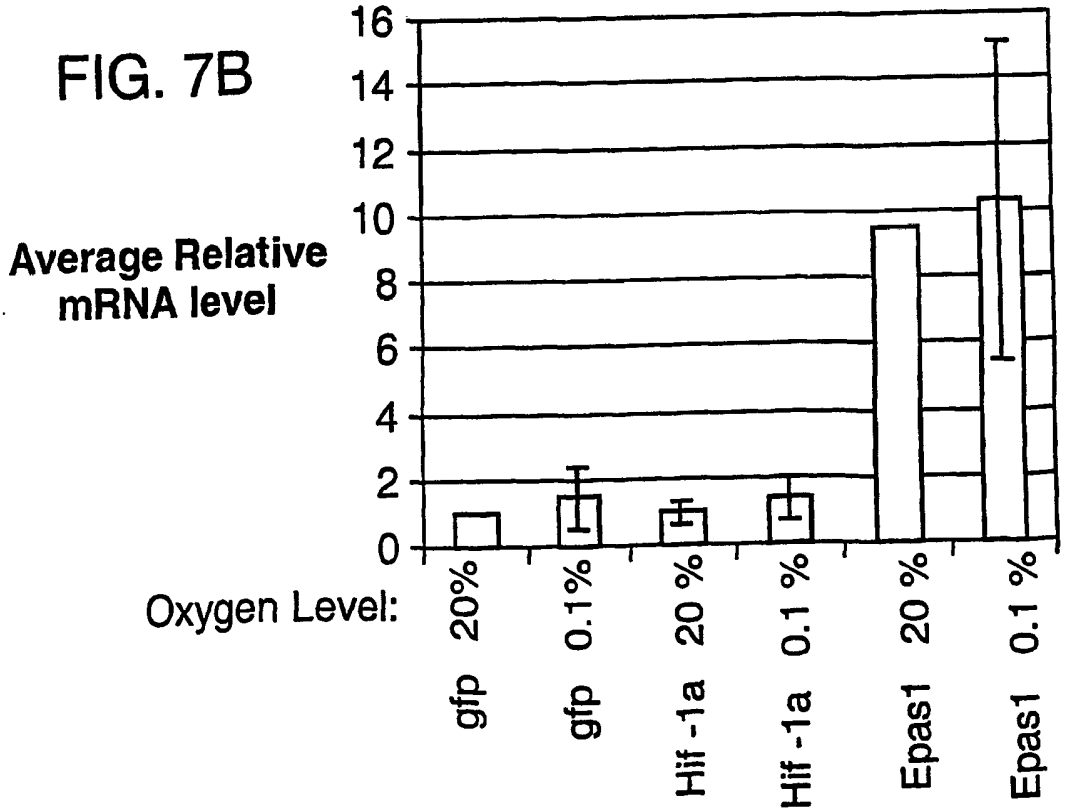


FIG. 7B



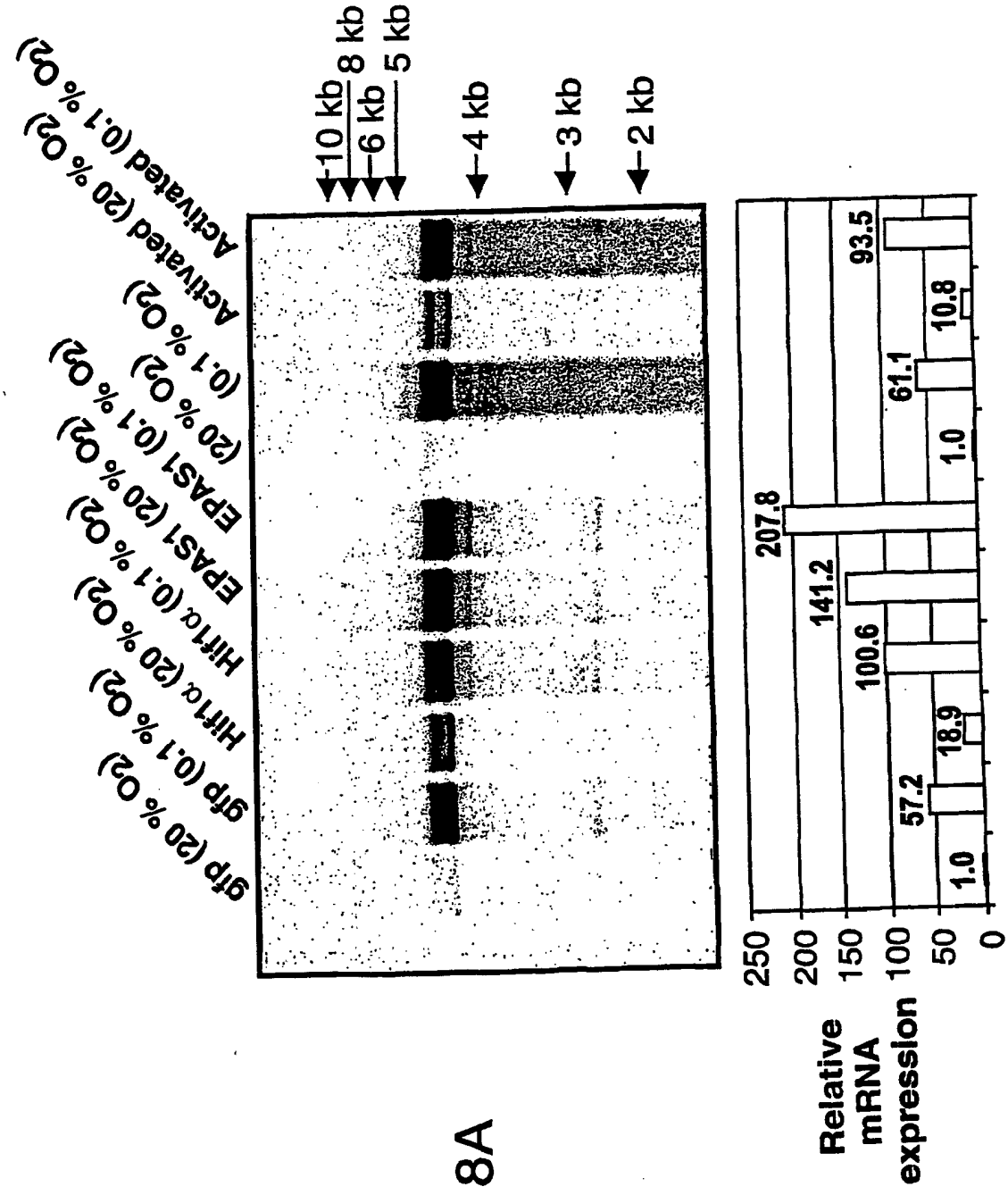

10 kb
8 kb
6 kb
5 kb
4 kb
3 kb
2 kb
Activated (0.1 % O2)
Activated (20 % O2)
(0.1 % O2)
(20 % O2)
EPAS1 (0.1 % O2)
EPAS1 (20 % O2)
Hif1α (0.1 % O2)
Hif1α (20 % O2)
gfp (0.1 % O2)
gfp (20 % O2)
Relative mRNA expression
250
200
150
100
50
0
207.8
141.2
100.6
57.2
18.9
1.0
93.5
61.1
10.8
1.0

FIG. 8A

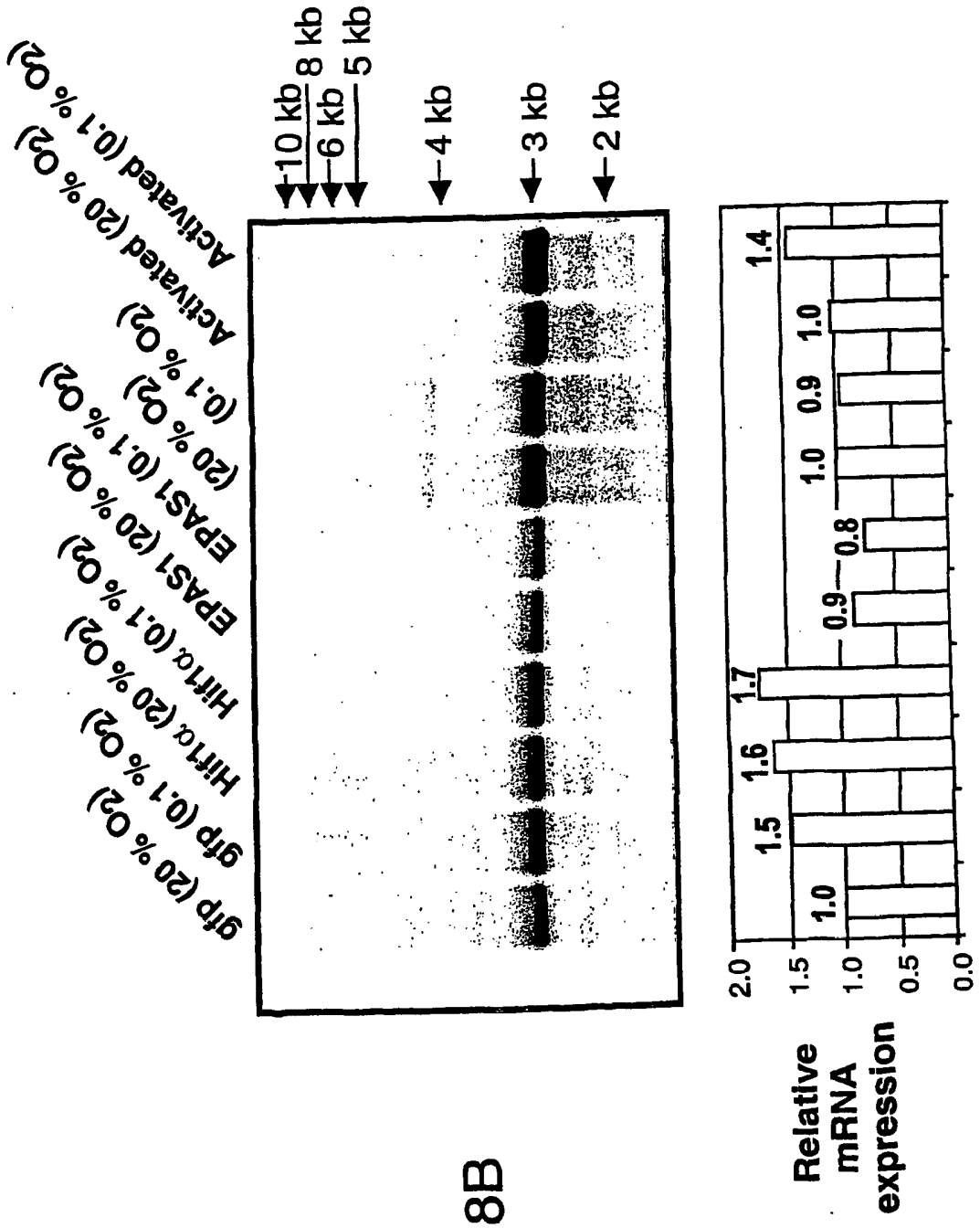

10 kb
8 kb
6 kb
5 kb
4 kb
3 kb
2 kb
Activated (0.1 % O2)
Activated (20 % O2)
(0.1 % O2)
(20 % O2)
EPAS1 (0.1 % O2)
EPAS1 (20 % O2)
Hif1α (0.1 % O2)
Hif1α (20 % O2)
gfp (0.1 % O2)
gfp (20 % O2)
1.4
1.0
0.9
1.0
0.8
0.9
1.7
1.6
1.5
1.0
2.0
1.5
1.0
0.5
0.0
Relative mRNA expression

FIG. 8B

FIG. 9

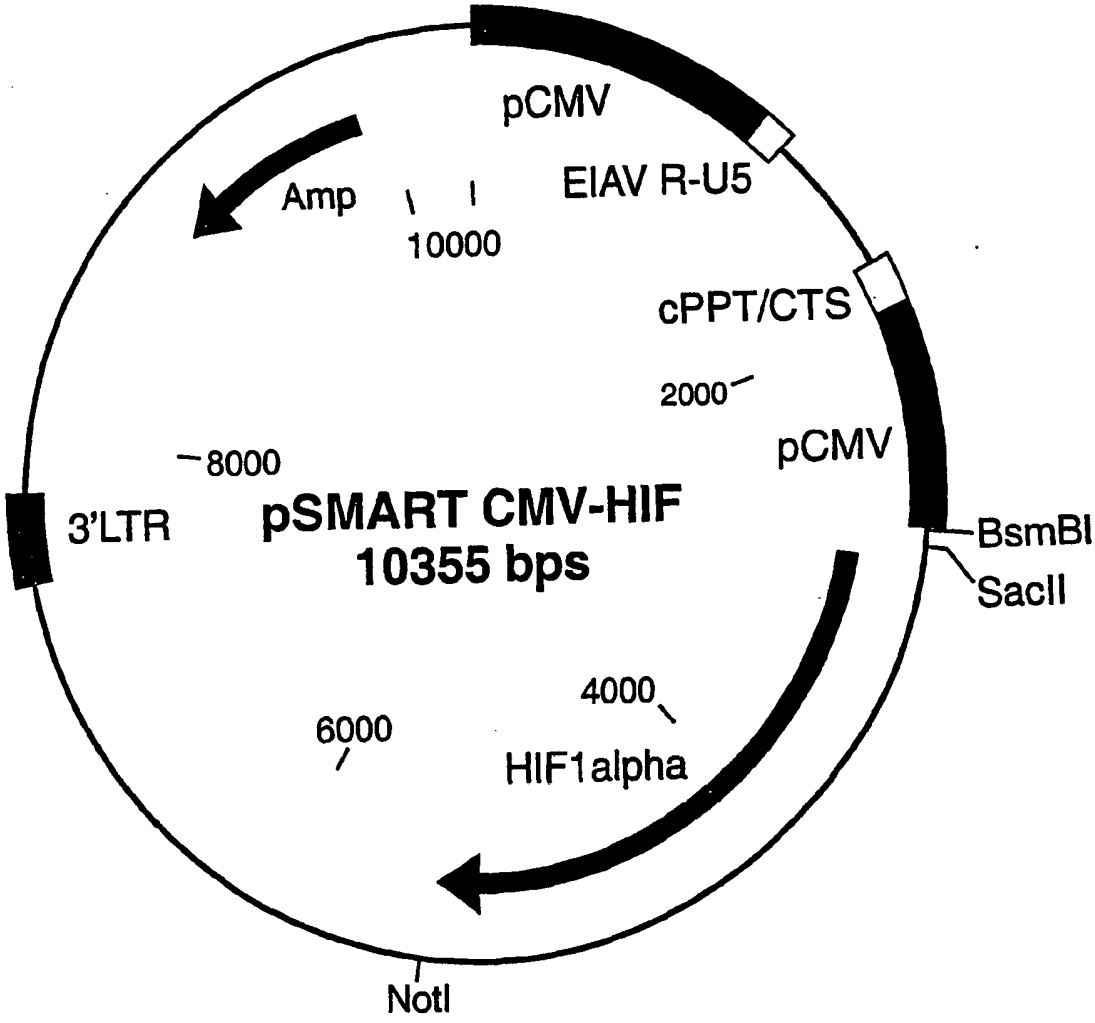
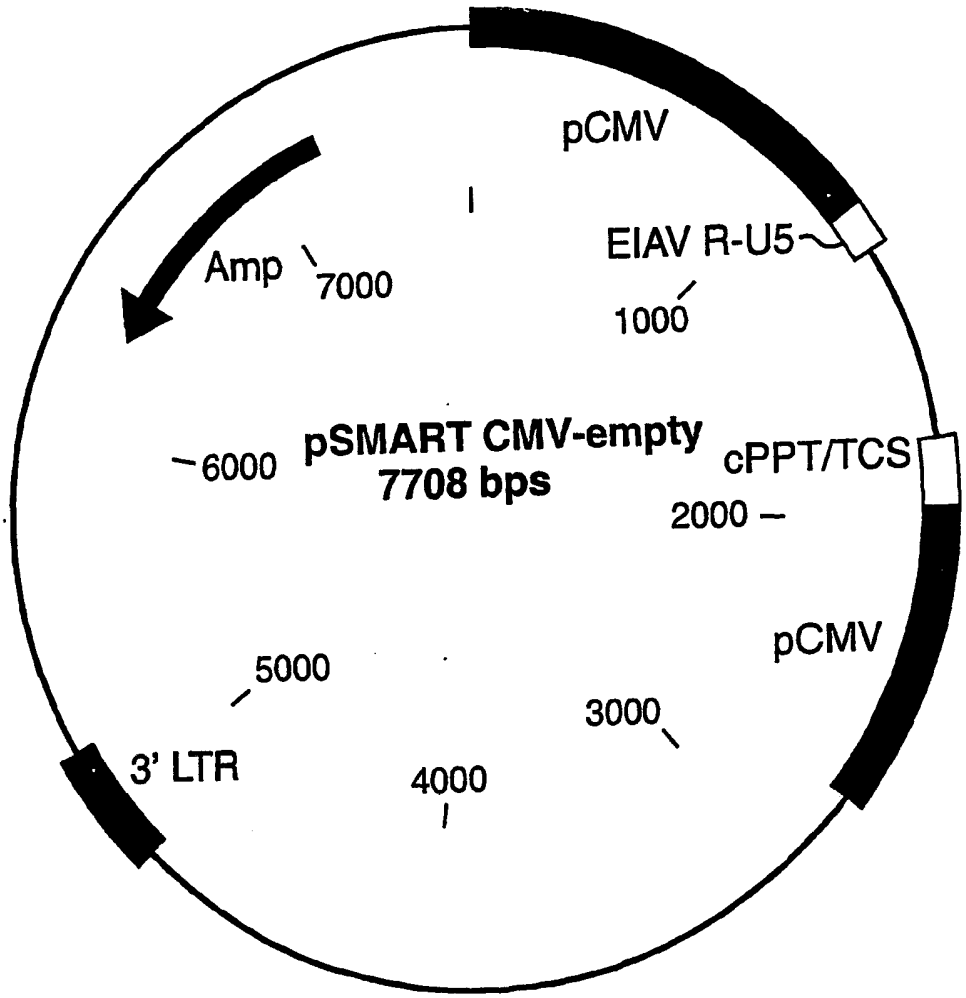

pCMV
EIAV R-U5
Amp
10000
cPPT/CTS
2000
pCMV
8000
3'LTR
pSMART CMV-HIF
10355 bps
BsmBI
SacII
4000
6000
HIF1alpha
NotI

FIG. 10


pCMV
EIAV R-U5
1000
cPPT/TCS
2000
pCMV
3000
4000
5000
3' LTR
6000
Amp
7000
pSMART CMV-empty
7708 bps

SEQUENCE LISTING

Seq ID 2

CAGACGGCGGAGAGCAGAGAGGGAGCGCGCCTTGGCTCGCTGGCCTTGGCGGCGGCTCCTCAGGAG
AGCTGGGGCGCCCACGAGAGGATCCCTCACCCGGGTCTCTCCTCAGGGATGACATCATCCGTCCAC
CTCCTTGTCTTCAAGGACCACCTCCTCTCCATGCTGAGCTGCTGCCAAGGGGCCTGCTGCCCATCT
ACACCTCACGAGGGCACTAGGAGCACGGTTTCCTGGATCCCACCAACATACAAAGCAGCCACTCAC
TGACCCCCAGGACCAGGATGGCAAAGGATGAAGAGGACCGGAACTGACCAGCCAGCTGTCCCTCTT
ACCTAAAGACTTAAACCAATGCCCTAGTGAGGGGGCATTGGGCATTAAGCCCTGACCTTTGCTATG
CTCATACTTTGACTCTATGAGTACTTTCCTATAAGTCTTTGCTTGTGTTCACCTGCTAGCAAACTG
GAGTGTTTCCCTCCCCAAGGGGGTGTCAGTCTTTGTCGACTGACTCTGTCATCACCCTTATGATGT
CCTGAATGGAAGGATCCCTTTGGAAATTCTCAGGAGGGGGACCTGGGCCAAGGGCTTGGCCAGCAT
CCTGCTGCAACTCCAAGGCCCTGGGTGGGCTTCTGGAATGAGCATGCTACTGAATCACCAAAGGCA
CGCCCCACCTCTCTGAAGATCTTCCTATCCTTTTCTGGGGGAATGGGGTCGATGAGAGCAACCTCC
TAGGGTTGTTGTGAGAATTAAATGAGATAAAAGAGGCCTCAGGCAGGATCTGGCATAGAGGAGGTG
ATCAGCAAATGTTTGTTGAAAAGGTTTGACAGGTCAGTCCCTTCCCACCCCTCTTGCTTGTCTTAC
TTGTCTTATTTATTCTCCAACAGCACTCCAGGCAGCCCTTGTCCACGGGCTCTCCTTGCATCAGCC
AAGCTTCTTGAAAGGCCTGTCTACACTTGCTGTCTTCCTTCCTCACCTCCAATTTCCTCTTCAACC
CACTGCTTCCTGACTCGCTCTACTCCGTGGAAGCACGCTCACAAAGGCACGTGGGCGTGGCCCGGC
TGGGTCGGCTGAAGAACTGCGGATGGAAGCTGCGGAAGAGCCCTGATGGGGCCCACCATCCCGGAC
CCAAGTCTTCTTCCTGGCGGGCCTCTCGTCTCCTTCCTGGTTTGGGCGGAAGCCATCACCTGGATG
CCTACGTGGGAAGGGACCTCGAATGTGGGACCCCAGCCCCTCTCCAGCTCGAAATCCCTCCACAGC
CACGGGGACACCCTGCACCTATTCCCACGGGACAGGCTGGACCCAGAGACTCTGGACCCGGGGCCT
CCCCTTGAGTAGAGACCCGCCCTCTGACTGATGGACGCCGTGACCTGGGGTCAGACCCGTGGGCTG
GACCCCTGCCCACCCCGCAGGAACCCTGAGGCCTAGGGGAGCTGTTGAGCCTTCAGTGTCTGCATG
TGGGAAGTGGGCTCCTTCACCTACCTCACAGGGCTGTTGTGAGGGGCGCTGTGATGCGGTTCCAAA
GCACAGGGCTTGGCGCACCCCACTGTGCTCTCAATAAATGTGTTTCCTGTCTTAACAAAAAAAAAA
AAAAAAAAAA

Seq ID 3

MELQTLQEALKVEIQVHQKLVAQMKQDPQNADLKKQLHELQAKITALSEKQKRVVEQLRKNLIVKQ
EQPDKFQIQPLPQSENKLQTAQQQPLQQLQQQQQQYHHHHAQQSAAASPNLTASQKTVTTASMITTK
TLPLVLKAATATMPASVVGQRPTIAMVTAINSQKAVLSTDVQNTPVNLQTSSKVTGPGAEAVQIVA
KNTVTLQVQATPPQPIKVPQFIPPPRLTPRPNFLPQVRPKPVAQNNIPIAPAPPPMLAAPQLIQRP
VMLTKFTPTTLPTSQNSIHPVRVVNGQTATIAKTFPMAQLTSIVIATPGTRLAGPQTVQLSKPSLE
KQTVKSHTETDEKQTESRTITPPAAPKPKREENPQKLAFMVSLGLVTHDHLEEIQSKRQERKRRTT
ANPVYSGAVFEPERKKSAVTYLNSTMHPGTRKRGRPPKYNAVLGFGALTPTSPQSSHPDSPENEKT
ETTFTFPAPVQPVSLPSPTSTDGDIHEDFCSVCRKSGQLLMCDTCSRVYHLDCLDPPLKTIPKGMW
ICPRCQDQMLKKEEAIPWPGTLAIVHSYIAYKAAKEEEKQKLLKWSSDLKQEREQLEQKVKQLSNS
ISKCMEMKNTILARQKEMHSSLEKVKQLIRLIHGIDLSKPVDSEATVGAISNGPDCTPPANAATST
PAPSPSSQSCTANCNQGEETK

Seq ID 4

GTGAGAGAGAGAGGAGAGGAGAGGAGGAGGAGGAGGAGGGAGAAGGGAACAACCTACCATCTTAAC
ACACTAATATCTAAAAAGTGCGAGAGGCCCAGAGCAGCAGCAGAAGCAGCAGCAGCAGCTCCAGCT
TCTTCCCTCCCTCCCCATGAAGAAGAGTTCCCTCCTCCTCCTCCTCCTGCTTCTCCTGCTCAGAGT
TCCTGCCTCCAGCTGCCAGGGGGGACAGCCAGCCAGCAGCAGGAGGGGGGCTAGAGAGCTGAAGGA
GAGCCAGTTTCCCCAAAATTGCTGCAGTGAGAAGAGGAGTTTGTTACTTTAAACAGAGGCTGAAGA
AACTATAGAATTAGCAGAGAAAGTGGAGAAGGTAGAGGATGGAGTTGCAGACTCTACAGGAGGCTC
TTAAAGTGGAAATTCAGGTTCACCAGAAACTGGTTGCTCAAATGAAGCAGGATCCACAGAATGCTG
ACTTAAAGAAACAGCTTCATGAACTCCAAGCCAAAATCACAGCTTTGAGTGAGAAACAGAAAAGAG
TAGTTGAACAGCTACGGAAGAACCTGATAGTAAAGCAAGAACAACCGGACAAGTTCCAAATACAGC
CATTGCCACAATCTGAAAACAAACTACAAACAGCACAGCAGCAACCACTACAGCAACTACAACAAC
AGCAGCAGTACCACCACCACCACGCCCAGCAGTCAGCTGCAGCCTCTCCCAACCTGACTGCTTCAC

AGAAGACTGTAACTACAGCTTCTATGATTACCACAAAGACACTACCTCTCGTCTTGAAAGCAGCAA
CTGCGACCATGCCTGCCTCTGTGGTGGGCCAGAGACCTACCATTGCTATGGTGACCGCCATCAACA
GTCAGAAGGCTGTGCTCAGCACTGATGTGCAGAACACACCAGTCAACCTCCAGACGTCTAGTAAGG
TCACTGGGCCTGGGGCAGAGGCTGTCCAAATTGTGGCAAAAAACACAGTCACTCTGCAGGTTCAGG
CAACACCTCCTCAGCCCATCAAAGTACCACAGTTTATCCCCCCTCCTAGACTCACTCCACGTCCAA
ACTTTCTTCCACAGGTTCGACCCAAGCCTGTGGCCCAGAATAACATTCCTATTGCCCCAGCACCAC
CTCCCATGCTCGCAGCTCCTCAGCTTATCCAGAGGCCCGTCATGCTGACCAAGTTCACCCCCACAA
CCCTTCCCACATCCCAGAATTCCATCCACCCCGTCCGTGTCGTCAATGGGCAGACTGCAACCATAG
CCAAAACGTTCCCCATGGCCCAGCTCACCAGCATTGTGATAGCTACTCCAGGGACCAGACTCGCTG
GACCTCAAACTGTACAGCTTAGCAAGCCAAGTCTTGAAAAACAGACAGTTAAATCTCACACAGAAA
CAGATGAGAAACAAACAGAGAGCCGCACCATCACCCCACCTGCTGCACCCAAACCAAAACGGGAGG
AGAACCCTCAGAAACTTGCCTTCATGGTGTCTCTAGGGTTGGTAACACATGACCATCTAGAAGAAA
TCCAAAGCAAGAGGCAAGAGCGAAAAAGAAGAACAACAGCAAATCCGGTCTACAGTGGAGCAGTCT
TTGAGCCAGAGCGTAAGAAGAGTGCAGTGACATACCTAAACAGCACAATGCACCCTGGGACCCGGA
AGAGAGGTCGTCCTCCAAAATACAATGCAGTGCTGGGGTTTGGAGCCCTTACCCCAACATCCCCCC
AATCCAGTCATCCTGACTCCCCTGAAAATGAAAAGACAGAGACCACATTCACTTTCCCTGCACCTG
TTCAGCCTGTGTCCCTGCCCAGCCCCACCTCCACAGACGGTGATATTCATGAGGATTTTTGCAGCG
TTTGCAGAAAAAGTGGCCAGTTACTGATGTGCGACACATGTTCCCGTGTATATCATTTGGACTGCT
TAGACCCCCCTCTGAAAACAATTCCCAAGGGCATGTGGATCTGTCCCAGATGTCAGGACCAGATGC
TGAAGAAGGAAGAAGCAATTCCATGGCCTGGAACTTTAGCAATTGTTCATTCCTATATTGCCTACA
AAGCAGCAAAAGAAGAAGAGAAACAGAAGTTACTTAAATGGAGTTCAGATTTAAAACAAGAACGAG
AACAACTAGAGCAAAAGGTGAAACAGCTCAGCAATTCCATAAGTAAATGCATGGAAATGAAGAACA
CCATCCTGGCCCGGCAGAAGGAGATGCACAGCTCCCTGGAGAAGGTAAAACAGCTGATTCGCCTCA
TCCACGGCATCGACCTCTCCAAACCTGTAGACTCTGAGGCCACTGTGGGGGCCATCTCCAATGGCC
CGGACTGCACCCCCCCTGCCAATGCCGCCACCTCCACGCCGGCCCCTTCCCCCTCCTCCCAGAGCT
GCACAGCGAACTGTAACCAGGGGGAAGAGACTAAATAACAGAGCCCCTCTAGGAGAAGCCACGGGA
TCCCGGCGGCAAGGAGAACAGAACACTGAAGACTCTAGAAAAGCAAAGCCGGATTTCTGGAAAGTG
CAGAATTCTTTTGGTTCTTTGGTTCCAGAGAGAGAGAAGATGCTTGTGCCAGGTGGCACCAGAGTT
TGCCAATTGATCCTTCTTATTCTGTGTGTACATGCAAAGATTGGACCATGTTACATGAAATAGTGC
CAGCTGGAGGTTCTTTGCCAGCACCATGCCAAGTGAAATAATATATTTACTCTCTCTATTATACAC
CAGTGTGTGCCTGCAGCAGCCTCCACAGCCACGATGGGTTTGTTTCTGTTTTCTTGGGTGGGGAGC
AGGGACGGGCGGAGGGAGGAGAGCAGGTTTCAGATCCTTACTTGCCGAGCCGTTTGTTTAGGTAGA
GAAGACAAGTCCAAAGAGTGTGTGGGCTTTCCTGTTTCTAAACTTTCGCTACTATAAAACCAAAAA
AAGGAATTGAGATTTCACCAACCCCAGTGCCCAGAAGAGGGAAGGGGAGTGGCTGGAGGGAGCAGG
GGGTGGGACAGTGTATCAAATAAGCAGTATTTAATCACCTCTGGCGGGGGCCTCGTGCAAGGGGAG
ACTGACACCAAGAACAGCCAGTAGGTTCTTCTCCCCTGCACTCTGCTCCCTGCGCGGTAACCCCAC
CACTCCTGAAGCCTGCCCAGTCTCCTTCCTTCCCTGCTTGGTGAGTCGCGCATCTCCGTGGTTATC
CCGCTGTCTCCTCTCCAAGAACAAGCAGAGCCCGGGCCACTGGCCCTTGCCCAAGGCAGGGAAGAA
GGATGTGTGTGTCCAGGAAGGAAAAAAAGGTGGATCAGTGATTTTACTTGAAAACAAGCTCCATCC
CTTTTCTATATTTATAAGAAGAGAAGATCTTGAGTGAAGCAGCACGCGACCCAGGTGTGTGTGAAT
TGAATGGAGACGTTTCTTTTCTCTTTCTTTAATTTTTGTTTTTGTTCTTTTTTTCTTTAAGGAAAG
TTTTATTTTACTGTTCATTTTACTTTCTTGGTAACAAAAACTAAAATAAGGAATAGAAAAGCTGTT
TTTCAGGCTGACAGTCCAATTAAGGGTAGCCAAGACCTTGCATGGTAGAGTAGGAATCATAGTGTC
AGTGAGGTCCCGTGAGTCTTTGTGAGTCCTTGTGTCATCGTTCGGGCACTGTTTTTTTATGCAAGG
GCAAAAATCTTTGTATCTGGGGAAAAAAAACTTTTTTTTAAATTAAAAAGGAAAATAAAAGATATT
GAGGTCTTCCTAGTGTTACTTAAATTAAGATCAAGGTAAGAAACATTGTAAAAAAAAATTACAAAA
GTGCTATTTGTTTCCTAAAAACAGTGATTTCTATTAAAAAGGTGTCAGAACTGG

Seq ID 5

MPLGHIMRLDLEKIALEYIVPCLHEVGFCYLDNFLGEVVGDCVLERVKQLHCTGALRDGQLAGPRA
GVSKRHLRGDQITWIGGNEEGCEAISFLLSLIDRLVLYCGSRLGKYYVKERSKAMVACYPGNGTGY
VRHVDNPNGDGRCITCIYYLNKNWDAKLHGGILRIFPEGKSFIADVEPIFDRLLFFWSDRRNPHEV
QPSYATRYAMTVWYFDAEERAEAKKKFRNLTRKTESALTED

Seq ID 6

GAGTCTGGCCGCAGTCGCGGCAGTGGTGGCTTCCCATCCCCAAAAGGCGCCCTCCGACTCCTTGCG
CCGCACTGCTCGCCGGGCCAGTCCGGAAACGGGTCGTGGAGCTCCGCACCACTCCCGCTGGTTCCC
GAAGGCAGATCCCTTCTCCCGAGAGTTGCGAGAAACTTTCCCTTGTCCCCGACGCTGCAGCGGCTC
GGGTACCGTGGCAGCCGCAGGTTTCTGAACCCCGGGCCACGCTCCCCGCGCCTCGGCTTCGCGCTC
GTGTAGATCGTTCCCTCTCTGGTTGCACGCTGGGGATCCCGGACCTCGATTCTGCGGGCGAGATGC
CCCTGGGACACATCATGAGGCTGGACCTGGAGAAAATTGCCCTGGAGTACATCGTGCCCTGTCTGC
ACGAGGTGGGCTTCTGCTACCTGGACAACTTCCTGGGCGAGGTGGTGGGCGACTGCGTCCTGGAGC
GCGTCAAGCAGCTGCACTGCACCGGGGCCCTGCGGGACGGCCAGCTGGCGGGGCCGCGCGCCGGCG
TCTCCAAGCGACACCTGCGGGGCGACCAGATCACGTGGATCGGGGGCAACGAGGAGGGCTGCGAGG
CCATCAGCTTCCTCCTGTCCCTCATCGACAGGCTGGTCCTCTACTGCGGGAGCCGGCTGGGCAAAT
ACTACGTCAAGGAGAGGTCTAAGGCAATGGTGGCTTGCTATCCGGGAAATGGAACAGGTTATGTTC
GCCACGTGGACAACCCCAACGGTGATGGTCGCTGCATCACCTGCATCTACTATCTGAACAAGAATT
GGGATGCCAAGCTACATGGTGGGATCCTGCGGATATTTCCAGAGGGGAAATCATTCATAGCAGATG
TGGAGCCCATTTTTGACAGACTCCTGTTCTTCTGGTCAGATCGTAGGAACCCACACGAAGTGCAGC
CCTCTTACGCAACCAGATATGCTATGACTGTCTGGTACTTTGATGCTGAAGAAAGGGCAGAAGCCA
AAAAGAAATTCAGGAATTTAACTAGGAAAACTGAATCTGCCCTCACTGAAGACTGACCGTGCTCTG
AAATCTGCTGGCCTTGTTCATTTTAGTAACGGTTCCTGAATTCTCTTAAATTCTTTGAGATCCAAA
GATGGCCTCTTCAGTGACAACAATCTCCCTGCTACTTCTTGCATCCTTCACATCCCTGTCTTGTGT
GTGGTACTTCATGTTTTCTTGCCAAGACTGTGTTGATCTTCAGATACTCTCTTTGCCAGATGAAGT
TATTTGCTAACTCCAGAAATTCCTGCAGACATCCTACTCGGCCAGCGGTTTACCTGATAGATTCGG
TAATACTATCAAGAGAAGAGCCTAGGAGCACAGCGAGGGAATGAACCTTACTTGCACTTTATGTAT
ACTTCCTGATTTGAAAGGAGGAGGTTTGAAAAGAAAAAAATGGAGGTGGTAGATGCCACAGAGAGG
CATCACGGAAGCCTTAACAGCAGGAAACAGAGAAATTTGTGTCATCTGAACAATTTCCAGATGTTC
TTAATCCAGGGCTGTTGGGGTTTCTGGAGAATTATCACAACCTAATGACATTAATACCTCTAGAAA
GGGCTGCTGTCATAGTGAACAATTTATAAGTGTCCCATGGGGCAGACACTCCTTTTTTCCCAGTCC
TGCAACCTGGATTTTCTGCCTCAGCTCCATTTTGCTGAAAATAATGACTTTCTGAATAAAGATGGC
AACACAATTTTTTCTCCATTTTCAGTTCTTACCTGGGAACCTAATTCCCCAGAAGCTAAAAAACTA
GACATTAGTTGTTTTGGTTGCTTTGTTGGAATGGAATTTAAATTTAAATGAAAGGAAAAATATATC
CCTGGTAGTTTTGTGTTAACCACTGATAACTGTGGAAAGAGCTAGGTCTACTGATATACAATAAAC
ATGTGTGCATCTTGAACAATTTGAGAGGGGAGGTGGAGTTGGAAATGTGGGTGTTCCTGTTTTTTT
TTTTTTTTTTTTTTTTTTTTAGTTTTCCTTTTTAATGAGCTCACCCTTTAACACAAAAAAAGCAGG
GTGATGTATTTTAAAAAAGGAAGTGGAAATAAAAAAATCTCAAAGCTATTTGAGTTCTCGTCTGTC
CCTAGCAGTCTTTCTTCAGCTCACTTGGCTCTCTAGATCCACTGTGGTTGGCAGTATGACCAGAAT
CATGGAACTTGCTAGAACTGTGGAAGCTTCTACTCCTGCAGTAAGCACAGATCGCACTGCCTCAAT
AACTTGGTATTGAGCACGTATTTTGCAAAAGCTACTTTTCCTAGTTTTCAGTATTACTTTCATGTT
TTAAAAATCCCTTTAATTTCTTGCTTGAAAATCCCATGAACATTAAAGAGCCAGAAATATTTTCCT
TTGTTATGTACGGATATATATATATATATAGTCTTCCAAGATAGAAGTTTACTTTTTCCTCTTCTG
GTTTTGGAAAATTTCCAGATAAGACATGTCACCATTAATTCTCAACGACTGCTCTATTTTGTTGTA
CGGTAATAGTTATCACCTTCTAAATTACTATGTAATTTACTCACTTATTATGTTTATTGTCTTGTA
TCCTTTCTCTGGAGTGTAAGCACAATGAAGACAGGAATTTTGTATATTTTTAACCAATGCAACATA
CTCTCAGCACCTAAAATAGTGCCGGGAACATAGTAAGGGCTCAGTAAATACTTGTTGAATAAACTC
AGTCTCCTACATTAGCATTCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAG

Seq ID 7

MACGDHFFVCVQRHTGFAFNDLLWLENTIKTNWLLDKLYITRSSLAAFSCTSVLYQIHKELFWEVR
STSAWPPCLGSEEPRFPGAAPSGK

Seq ID 8

CTCTTGACCAGAAGATTTGTTTACCTACCATTTCATTGGTAGTATGGCCTGCGGTGACCATTTTTT
TGTGTGTGTACAGCGTCATACAGGCTTTGCCTTTAATGATCTCTTATGGTTAGAAAACACAATAAA
AACAAACTGGCTACTGGACAAATTGTATATTACCAGATCATCACTAGCAGCTTTCAGTTGCACTTC
AGTCCTTTATCAAATTCATAAAGAATTGTTCTGGGAAGTGAGGAGCACCTCTGCCTGGCCACCCTG
TCTGGGAAGTGAGGAGCCCCGCTTCCCGGGAGCCGCCCCATCTGGGAAGTGAGGAGCGCTTGCCTG
GCCGCCCAACTGTCTGGGAAGTGAGGTGTGCCTGTGCCTGGCTGCCGGCTTCACTCCATCAAACAC
GGCTGGGAGTTCGGGAGTCTCCAGCATGGCATTCATCTTGATCAATGTTACTTATTCTTCCCACTG

CAAATACGAATTNTCCTTGAAGAGTTGTACAGNTGGGCCGGGCAGGTCAGATCTCTTATTCAAACC
AACCTAAAAGCTGCGTGTGATTGACTACAGGTGCTCAGCTCGGCCCCAGTCTGGTTTCTGCAGCTT
CCCG

Seq ID 10

TTAATGTCGAAAATGCAAACTTGGGGAGGGCAGAAAGATCACACACAAGGCTGTCACTTCATACTT
GCAGGATTGCACAGCAGCCGGGCAGAGGCGCTCCTCACTTCCCAGATGGGGCGGCGGGCANCAGAG
ACGCACCTCACTTCCTAGACAGTGCGGCANCATGGCACAGGCACACCTCACTTCCCAGACAGTTGG
GCGGCCAGGCAAGCGCTCCTCACTTCCCAGATGGGGCGGCTCCCGGGAAGCGGGGCTCCTCACTTC
CCAGACAGGGTGGCCAGGCAGAGGTGCTCCTCACTTCCCAGAACAATTCTTTATGAATTTGATAAA
GGACTGAAGTGCAACTGAAAGCTGCTAGTGATGATCTGGTAATATACAATTTGTCCAGTAGCCAGT
TTGTTTTTATTGTGTTTTCTAACCATAAGAGATCATTAAAGGCAAAGCTGTTATGACG

Seq ID 11

MLAVGCALLAALLAAPGAALAPRRCPAQEVARGVLTSLPGDSVTLTCPGVEPEDNATVHWVLRKPA
AGSHPSRWAGMGRRLLLRSVQLHDSGNYSCYRAGRPAGTVHLLVDVPPEEPQLSCFRKSPLSNVVC
EWGPRSTPSLTTKAVLLVRKFQNSPAEDFQEPCQYSQESQKFSCQLAVPEGDSSFYIVSMCVASSV
GSKFSKTQTFQGCGILQPDPPANITVTAVARNPRWLSVTWQDPHSWNSSFYRLRFELRYRAERSKT
FTTWMVKDLQHHCVIHDAWSGLRHVVQLRAQEEFGQGEWSEWSPEAMGTPWTESRSPPAENEVSTP
MQALTTNKDDDNILFRDSANATSLPVQDSSSVPLPTFLVAGGSLAFGTLLCIAIVLRFKKTWKLRA
LKEGKTSMHPPYSLGQLVPERPRPTPVLVPLISPPVSPSSLGSDNTSSHNRPDARDPRSPYDISNT
DYFFPR

Seq ID 12

GGCGGTCCCCTGTTCTCCCCGCTCAGGTGCGGCGCTGTGGCAGGAAGCCACCCCCTCGGTCGGCCG
GTGCGCGGGGCTGTTGCGCCATCCGCTCCGGCTTTCGTAACCGCACCCTGGGACGGCCCAGAGACG
CTCCAGCGCGAGTTCCTCAAATGTTTTCCTGCGTTGCCAGGACCGTCCGCCGCTCTGAGTCATGTG
CGAGTGGGAAGTCGCACTGACACTGAGCCGGGCCAGAGGGAGAGGAGCCGAGCGCGGCGCGGGGCC
GAGGGACTCGCAGTGTGTGTAGAGAGCCGGGCTCCTGCGGATGGGGGCTGCCCCCGGGGCCTGAGC
CCGCCTGCCCGCCCACCGCCCCGCCCCGCCCCTGCCACCCCTGCCGCCCGGTTCCCATTAGCCTGT
CCGCCTCTGCGGGACCATGGAGTGGTAGCCGAGGAGGAAGCATGCTGGCCGTCGGCTGCGCGCTGC
TGGCTGCCCTGCTGGCCGCGCCGGGAGCGGCGCTGGCCCCAAGGCGCTGCCCTGCGCAGGAGGTGG
CAAGAGGCGTGCTGACCAGTCTGCCAGGAGACAGCGTGACTCTGACCTGCCCGGGGGTAGAGCCGG
AAGACAATGCCACTGTTCACTGGGTGCTCAGGAAGCCGGCTGCAGGCTCCCACCCCAGCAGATGGG
CTGGCATGGGAAGGAGGCTGCTGCTGAGGTCGGTGCAGCTCCACGACTCTGGAAACTATTCATGCT
ACCGGGCCGGCCGCCCAGCTGGGACTGTGCACTTGCTGGTGGATGTTCCCCCCGAGGAGCCCCAGC
TCTCCTGCTTCCGGAAGAGCCCCCTCAGCAATGTTGTTTGTGAGTGGGGTCCTCGGAGCACCCCAT
CCCTGACGACAAAGGCTGTGCTCTTGGTGAGGAAGTTTCAGAACAGTCCGGCCGAAGACTTCCAGG
AGCCGTGCCAGTATTCCCAGGAGTCCCAGAAGTTCTCCTGCCAGTTAGCAGTCCCGGAGGGAGACA
GCTCTTTCTACATAGTGTCCATGTGCGTCGCCAGTAGTGTCGGGAGCAAGTTCAGCAAAACTCAAA
CCTTTCAGGGTTGTGGAATCTTGCAGCCTGATCCGCCTGCCAACATCACAGTCACTGCCGTGGCCA
GAAACCCCCGCTGGCTCAGTGTCACCTGGCAAGACCCCCACTCCTGGAACTCATCTTTCTACAGAC
TACGGTTTGAGCTCAGATATCGGGCTGAACGGTCAAAGACATTCACAACATGGATGGTCAAGGACC
TCCAGCATCACTGTGTCATCCACGACGCCTGGAGCGGCCTGAGGCACGTGGTGCAGCTTCGTGCCC
AGGAGGAGTTCGGGCAAGGCGAGTGGAGCGAGTGGAGCCCGGAGGCCATGGGCACGCCTTGGACAG
AATCCAGGAGTCCTCCAGCTGAGAACGAGGTGTCCACCCCCATGCAGGCACTTACTACTAATAAAG
ACGATGATAATATTCTCTTCAGAGATTCTGCAAATGCGACAAGCCTCCCAGTGCAAGATTCTTCTT
CAGTACCACTGCCCACATTCCTGGTTGCTGGAGGGAGCCTGGCCTTCGGAACGCTCCTCTGCATTG
CCATTGTTCTGAGGTTCAAGAAGACGTGGAAGCTGCGGGCTCTGAAGGAAGGCAAGACAAGCATGC
ATCCGCCGTACTCTTTGGGGCAGCTGGTCCCGGAGAGGCCTCGACCCACCCCAGTGCTTGTTCCTC
TCATCTCCCCACCGGTGTCCCCCAGCAGCCTGGGGTCTGACAATACCTCGAGCCACAACCGACCAG
ATGCCAGGGACCCACGGAGCCCTTATGACATCAGCAATACAGACTACTTCTTCCCCAGATAGCTGG
CTGGGTGGCACCAGCAGCCTGGACCCTGTGGATGACAAAACACAAACGGGCTCAGCAAAAGATGCT
TCTCACTGCCATGCCAGCTTATCTCAGGGGTGTGCGGCCTTTGGCTTCACGGAAGAGCCTTGCGGA
AGGTTCTACGCCAGGGGAAAATCAGCCTGCTCCAGCTGTTCAGCTGGTTGAGGTTTCAAACCTCCC

TTTCCAAATGCCCAGCTTAAAGGGGTTAGAGTGAACTTGGGCCACTGTGAAGAGAACCATATCAAG
ACTCTTTGGACACTCACACGGACACTCAAAAGCTGGGCAGGTTGGTGGGGGCCTCGGTGTGGAGAA
GCGGCTGGCAGCCCACCCCTCAACACCTCTGCACAAGCTGCACCCTCAGGCAGGTGGGATGGATTT
CCAGCCAAAGCCTCCTCCAGCCGCCATGCTCCTGGCCCACTGCATCGTTTCATCTTCCAACTCAAA
CTCTTAAAACCCAAGTGCCCTTAGCAAATTCTGTTTTTCTAGGCCTGGGGACGGCTTTTACTTAAA
CGCCAAGGCCTGGGGGAAGAAGCTCTCTCCTCCCTTTCTTCCCTACAGTTCAAAAACAGCTGAGGG
TGAGTGGGTGAATAATACAGTATGTCAGGGCCTGGTCGTTTTCAACAGAATTATAATTAGTTCCTC
ATTAGCAGTTTTGCCTAAATGTGAATGATGATCCTAGGCATTTGCTGAATACAGAGGCAACTGCAT
TGGCTTTGGGTTGCAGGACCTCAGGTGAGAAGCAGAGGAAGGAGAGGAGAGGGGCACAGGGTCTCT
ACCATCCCCTGTAGAGTGGGAGCTGAGTGGGGGATCACAGCCTCTGAAAACCAATGTTCTCTCTTC
TCCACCTCCCACAAAGGAGAGCTAGCAGCAGGGAGGGCTTCTGCCATTTCTGAGATCAAAACGGTT
TTACTGCAGCTTTGTTTGTTGTCAGCTGAACCTGGGTAACTAGGGAAGATAATATTAAGGAAGACA
ATGTGAAAAGAAAAATGAGCCTGGCAAGAATGCGTTTAAACTTGGTTTTTAAAAAACTGCTGACTG
TTTTCTCTTGAGAGGGTGGAATATCCAATATTCGCTGTGTCAGCATAGAAGTAACTTACTTAGGTG
TGGGGGAAGCACCATAACTTTGTTTAGCCCAAAACCAAGTCAAGTGAAAAAGGAGGAAGAGAAAAA
ATATTTTCCTGCCAGGCATGGAGGCCCACGCACTTCGGGAGGTCGAGGCAGGAGGATCACTTGAGT
CCAGAAGTTTGAGATCAGCCTGGGCAATGTGATAAAACCCCATCTCTACAAAAAGCATAAAAATTA
GCCAAGTGTGGTAGAGTGTGCCTGAAGTCCCAGATACTTGGGGGGCTGAGGTGGGAGGATCTCTTG
AGCCTGGGAGGTCAAGGCTGCAGTGAGCCGAGATTGCACCACTGCACTCCAGCCTGGGGTGACAGA
GCAAGTGAGACCCTGTCTC

Seq ID 13

MEGLVMFQDLSIDFSQEEWECLDAAQKDLYRDVMMENYSSLVSLGLSIPKPDVISLLEQGKEPWMV
SRDVLGGWCRDVFISTQQGNAMFLLSGYFSDSEFRCKTKDSCLPKEIYEVTSSQWVRMEKCHSLVG
SSVRDDWECKGQFQHQDINQERYLEKAIMTYETTPTFCLQTSLTLHHRIHPGEKLYKSTECMAFKY
GSELTQQQETHTGEKLYKCKECGKAFHHFSYLVKHQRIHTGEKPCACKEYGKAFISGSHLIQHQKM
YTDERPHECQESVKAFRPSAHLIQHWRIHTGDKPYECKECGKSFTSGSTLNQHQQIHTGEKPYHCK
QCGKSFTVGSTLIRHQQIHTGEKPYDCKECGKSFASGSALIRHQRIHTGEKPYDCKECGKSFTFHS
ALIRHQRIHTGEKPYDCKECGKSFTFRSGLIGHQAIHTGEKPYDCKECGKSFTAGSTLIQHQRIHT
GEKPYDCKECGKSFASGSALLQHQRIHTGEKPYCCKECGKSFTFRSTRNRHQRIHTGEKPYNCKEC
GKSFASGSALLQHQRIHTGEKPYHCKECGKSFTFRSGLIGHQAVHTGEKPYDCKECGKSFTSRSAL
IQHQRIHTGEKPYHCKECGKSFTVGSTLLQHQQIHTGEKPYDCKECGKAFRLRLRLTQHQQIHTGE
KPYQCQECGKAFVSVSGLTQHHRIHTGEKPYECPDCGKAFRQRTYLNQHRRIHTGEKPYECKECGK
SFTFCSGLIQHQQNHTDEKPYDGKECGKSFTSHSTLIQHQQIHTGEKPYDCKECGKSFTSHSTLIQ
HQQIHTGEKLYDCKECGKSFTSHSTLIQHQPLHTGEKPYHCKECGKSFTLRSALIQHRPVHTGEKR
YSCKECGKSFTSRSTLIEHQRIHTGEKPYHCKECGKSFAFRSAIIQHRRIHTGEKPYDCKECGKAF
RRRSKLTQHQRIHTGEKPYRCHECGKAFVRFSGLTKHHSIHTGEKPYECKTCGKSFRQRTHLTLHQ
RIHTGDRPYECKECGKSFTCGSELIRHQRTHTGEKPYDCKECGKAFRCPSQLSQHKRIHTGEKTYQ
CPECGKAFFYASGLSRHQSVHTGEKPYECKTCGKAFKQLTQLTRHQRIHDLT

Seq ID 14

ATGGAGGGGTTGGTCATGTTCCAGGATCTGTCTATAGACTTCTCTCAGGAGGAATGGGAGTGCCTG
GACGCTGCTCAGAAGGACTTATACAGAGATGTAATGATGGAGAACTACAGCAGCCTGGTCTCACTA
GGTCTCTCTATCCCAAAGCCTGATGTGATTTCCTTACTGGAGCAAGGGAAAGAGCCCTGGATGGTT
TCAAGGGACGTGCTGGGAGGATGGTGCCGAGATGTTTTTATTTCTACTCAACAAGGAAATGCTATG
TTTTTGCTTTCTGGTTATTTTTCAGACTCGGAGTTTAGATGTAAGACCAAAGATTCATGTTTGCCA
AAAGAAATCTATGAAGTAACATCATCTCAGTGGGTGAGAATGGAAAAATGTCATAGCCTTGTGGGC
TCCAGTGTCAGAGATGACTGGGAATGCAAAGGCCAGTTTCAGCACCAAGATATAAATCAGGAGCGA
TATTTGGAAAAAGCAATAATGACTTATGAAACAACACCAACTTTCTGCCTACAGACATCTCTCACT
CTGCATCATCGGATTCATCCTGGAGAGAAACTGTATAAATCCACAGAATGTATGGCTTTTAAGTAT
GGCTCAGAACTTACACAACAGCAAGAAACCCATACTGGTGAAAAACTCTATAAATGTAAGGAGTGT
GGGAAGGCCTTTCATCACTTTTCCTATCTTGTTAAACATCAGAGAATTCATACTGGGGAAAAGCCC
TGTGCATGTAAAGAATATGGAAAAGCTTTTATTTCTGGCTCACATCTTATTCAACACCAGAAAATG
TATACAGATGAGAGACCTCATGAGTGTCAGGAATCCGTGAAGGCCTTTAGACCGTCTGCACATCTT
ATTCAACATTGGAGAATTCATACTGGTGACAAACCTTATGAATGTAAAGAATGTGGGAAATCTTTT

ACTTCTGGCTCAACACTAAATCAACATCAGCAAATTCACACTGGTGAGAAACCCTATCATTGTAAG
CAATGTGGAAAATCTTTTACTGTTGGCTCAACACTAATTCGACACCAGCAAATTCACACTGGTGAG
AAACCGTATGATTGTAAGGAATGTGGAAAATCTTTTGCTTCAGGCTCAGCACTAATTCGACATCAG
CGAATTCACACTGGTGAGAAACCCTATGACTGTAAGGAATGTGGAAAATCTTTTACTTTTCACTCA
GCTCTAATTCGACATCAGCGAATTCACACTGGTGAGAAACCCTATGATTGTAAGGAATGTGGGAAA
TCTTTTACTTTTCGCTCAGGGTTAATTGGACACCAGGCAATTCACACTGGTGAGAAACCCTATGAT
TGTAAAGAATGTGGAAAATCTTTTACTGCTGGCTCAACACTAATTCAACATCAGCGAATTCACACT
GGTGAGAAACCCTATGATTGTAAGGAGTGTGGGAAATCTTTTGCTTCGGGCTCAGCACTACTTCAA
CATCAGCGGATTCACACTGGTGAGAAACCCTATTGTTGTAAGGAATGTGGAAAATCTTTTACTTTT
CGCTCAACACGCAATCGACACCAGCGAATCCACACTGGTGAGAAACCCTATAATTGTAAAGAATGT
GGAAAATCTTTTGCTTCTGGCTCAGCACTACTTCAACATCAGCGAATTCACACTGGTGAGAAACCC
TATCATTGTAAGGAATGTGGAAAATCTTTTACTTTTCGCTCAGGGCTAATTGGACATCAGGCAGTT
CACACTGGTGAGAAACCCTATGATTGTAAAGAATGTGGAAAATCTTTTACTTCTCGCTCAGCACTA
ATTCAACATCAGCGAATTCACACTGGTGAGAAACCCTATCACTGTAAGGAATGTGGAAAATCTTTT
ACTGTTGGCTCAACACTACTTCAACATCAGCAAATTCACACTGGTGAGAAACCTTATGATTGTAAG
GAATGTGGGAAGGCCTTTAGACTTCGTTTACGACTTACTCAACATCAACAAATCCATACTGGTGAG
AAACCTTATCAATGTCAGGAATGTGGGAAAGCCTTTGTCAGTGTCTCAGGACTCACCCAACATCAC
AGAATTCACACTGGTGAGAAACCCTATGAATGTCCGGACTGTGGGAAGGCCTTTAGACAGCGTACA
TACCTTAATCAACATCGGAGAATTCATACTGGTGAGAAACCTTATGAATGTAAAGAATGTGGGAAA
TCTTTTACTTTCTGCTCAGGACTAATTCAACATCAGCAAAATCACACTGATGAGAAACCCTATGAT
GGTAAGGAATGTGGGAAATCTTTTACTTCTCACTCAACACTAATTCAACATCAGCAAATTCACACT
GGTGAGAAACCCTATGATTGTAAGGAATGTGGGAAATCTTTTACTTCTCACTCAACACTAATTCAA
CATCAGCAAATACACACTGGTGAGAAGCTCTATGATTGTAAGGAATGTGGGAAATCTTTTACTTCT
CATTCAACACTAATTCAACATCAGCCACTTCACACTGGTGAGAAACCTTATCATTGCAAGGAATGT
GGGAAATCTTTTACTCTTCGCTCAGCACTAATTCAACATCGGCCAGTTCACACTGGTGAGAAACGC
TACAGTTGTAAAGAATGTGGGAAATCTTTTACTTCTCGCTCAACACTAATTGAACATCAGCGAATT
CACACTGGTGAGAAACCCTATCATTGTAAGGAATGTGGAAAATCTTTTGCTTTTCGCTCAGCAATA
ATCCAACATCGGCGAATTCACACTGGTGAGAAACCCTATGATTGTAAGGAATGTGGCAAGGCCTTT
AGACGTCGTTCAAAACTTACTCAACATCAACGAATCCATACTGGTGAGAAACCTTATCGATGTCAT
GAATGTGGAAAAGCCTTTGTCCGTTTCTCAGGACTCACCAAACATCACAGTATTCACACTGGCGAG
AAACCCTATGAATGTAAGACATGTGGGAAGTCCTTCAGACAGCGTACACACCTTACTCTACATCAG
AGAATTCATACCGGTGACAGACCTTATGAATGTAAAGAATGTGGGAAATCTTTTACTTGCGGCTCA
GAACTAATTCGACATCAGCGAACTCACACTGGTGAGAAACCTTATGATTGTAAGGAATGTGGGAAG
GCCTTTAGATGTCCTTCACAACTTAGTCAACATAAACGAATCCATACTGGTGAGAAAACTTATCAA
TGTCCGGAATGTGGGAAAGCCTTTTTCTATGCCTCAGGACTCAGCCGACATCAGAGTGTTCACACT
GGCGAGAAACCCTATGAATGTAAGACATGTGGGAAGGCCTTTAAACAGCTTACACAGCTTACTCGA
CATCAGAGAATTCATGACCTAACATAA

Seq ID 18

GATCTTCAGTTCCTATTTTACTATAAGCCTTTGAAATGTTAAGGCTGAATAGACCAAGATATGAAA
AGGGACAAGAGGAAGTAAGAACAGTGCAGATACAGTGGGAACACACGGTAACTTTCTGCCTTCTCC
ATTCTACGTGGACGGTGTGCTTTCCGGGCAGCAGGTGTTGTGACTGTAAAAAGATCACACATTTGT
CAGTAACTATGTAGGATTTTTGGCACTGTTAAGAGCGTANATNTATAAAGTCAGGGCAGCTAATAA
AAATTCCCTCAACCAAATTTTTGATAAGTCTGAAATTAAAAAAAAATAATTTTTCTACCAAGGCAA
ATTCTTGGGTTCTCCNTTTCAGCCCTGCCTTTTAGGAAAACTTACATCCTTTAAAAAACTTTTTAA
AAAAAGGNTTCCNTTCACNGCAGGNGGTGGACTTNGGG

Seq ID 19

MYQDYPGNFDTSSRGSSGSPAHAESYSSGGGGQQKFRVDMPGSGSAFIPTINAITTSQDLQWMVQP
TVITSMSNPYPRSHPYSPLPGLASVPGHMALPRPGVIKTIGTTVGRRRRDEQLSPEEEEKRRIRRE
RNKLAAAKCRNRRRELTEKLQAETEELEEEKSGLQKEIAELQKEKEKLEFMLVAHGPVCKISPEER
RSPPAPGLQPMRSGGGSVGAVVVKQEPLEEDSPSSSSAGLDKAQRSVIKPISIAGGFYGEEPLHTP
IVVTSTPAVTPGTSNLVFTYPSVLEQESPASPSESCSKAHRRSSSSGDQSSDSLNSPTLLAL

Seq ID 20

GGCGCTCGCAGAGCCGGAAAGAAGTGCTGTAAGGGACGCTCGGGGGACGCTGTTCCTGAGGTGTCG
CCGCCTCCCTGTCCTCGCCCTCCGCGGTGGGGGAGAAACCCAGGAGCGAAGCCCAGAGCCCGCGGC
GCGGCCGGCGGACGAACGAGCGCGCAGCAGCCGGTGCGCGGCCGCGGCGAGGGCGGGGGAAGAAAA
ACACCCTGTTTCCTCTCCGGCCCCCACCGCGGATCATGTACCAGGATTATCCCGGGAACTTTGACA
CCTCGTCCCGGGGCAGCAGCGGCTCTCCTGCGCACGCCGAGTCCTACTCCAGCGGCGGCGGCGGCC
AGCAGAAATTCCGGGTAGATATGCCTGGCTCAGGCAGTGCATTCATCCCCACCATCAACGCCATCA
CGACCAGCCAGGACCTGCAGTGGATGGTGCAGCCCACAGTGATCACCTCCATGTCCAACCCATACC
CTCGCTCGCACCCCTACAGCCCCCTGCCGGGCCTGGCCTCTGTCCCTGGACACATGGCCCTCCCAA
GACCTGGCGTGATCAAGACCATTGGCACCACCGTGGGCCGCAGGAGGAGAGATGAGCAGCTGTCTC
CTGAAGAGGAGGAGAAGCGTCGCATCCGGCGGGAGAGGAACAAGCTGGCTGCAGCCAAGTGCCGGA
ACCGACGCCGGGAGCTGACAGAGAAGCTGCAGGCGGAGACAGAGGAGCTGGAGGAGGAGAAGTCAG
GCCTGCAGAAGGAGATTGCTGAGCTGCAGAAGGAGAAGGAGAAGCTGGAGTTCATGTTGGTGGCTC
ACGGCCCAGTGTGCAAGATTAGCCCCGAGGAGCGCCGATCGCCCCCAGCCCCTGGGCTGCAGCCCA
TGCGCAGTGGGGGTGGCTCGGTGGGCGCTGTAGTGGTGAAACAGGAGCCCCTGGAAGAGGACAGCC
CCTCGTCCTCGTCGGCGGGGCTGGACAAGGCCCAGCGCTCTGTCATCAAGCCCATCAGCATTGCTG
GGGGCTTCTACGGTGAGGAGCCCCTGCACACCCCCATCGTGGTGACCTCCACACCTGCTGTCACTC
CGGGCACCTCGAACCTCGTCTTCACCTATCCTAGCGTCCTGGAGCAGGAGTCACCCGCATCTCCCT
CCGAATCCTGCTCCAAGGCTCACCGCAGAAGCAGTAGCAGCGGGGACCAATCATCAGACTCCTTGA
ACTCCCCCACTCTGCTGGCTCTGTAACCCAGTGCACCTCCCTCCCCAGCTCCGGAGGGGGTCCTCC
TCGCTCCTCCTTCCCAGGGACCAGCACCTTCAAGCGCTCCAGGGCCGTGAGGGCAAGAGGGGGACC
TGCCACCAGGGAGCTTCCTGGCTCTGGGGGACCCAGGTGGGACTTAGCAGTGAGTATTGGAAGACT
TGGGTTGATCTCTTAGAAGCCATGGGACCTCCTCCCTCATTCATCTTGCAAGCAAATCCCATTTCT
TGAAAAGCCTTGGAGAACTCGGTTTGGTAGACTTGGACATCTCTCTGGCTTCTGAAGAGCCTGAAG
CTGGCCTGGACCATTCCTGTCCCTTTGTTACCATACTGTCTCTGGAGTGATGGTGTCCTTCCCTGC
CCCACCACGCATGCTCAGTGCCTTTTGGTTTCACCTTCCCTCGACTTGACCCTTTCCTCCCCCAGC
GTCAGTTTCACTCCCTCTTGGTTTTTATCAAATTTGCCATGACATTTCATCTGGGTGGTCTGAATA
TTAAAGCTCTTCATTTCTGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Seq ID 21

MMKKSADHKNLEVIVTNGYDKDGFVQDIQNDIHASSSLNGRSTEEVKPIDENLGQTGKSAVCIHQD
INDDHVEDVTGIQHLTSDSDSEVYCDSMEQFGQEESLDSFTSNNGPFQYYLGGHSSQPMENSGFRE
DIQVPPGNGNIGNMQVVAVEGKGEVKHGGEDGRNNSGAPHREKRGGETDEFSNVRRGRGHRMQHLS
EGTKGRQVGSGGDGERWGSDRGSRGSLNEQIALVLMRLQEDMQNVLQRLQKLETLTALQAKSSTST
LQTAPQPTSQRPSWWPFEMSPGVLTFAIIWPFIAQWLVYLYYQRRRRKLN

Seq ID 22

cgctccatgtatnagtttcatgcaggctcttgggaaagctggtgctgctgctgcctgattcccgcc
gacagaccttgggaccggggccaacactggcagctggagatggcggacacgagatccgtgcacgag
actaggtttgaggcggccgtgaaggtgatccagagtttgccgaagaatgattcattccagccaaca
aatgaaatgatgcttaaattttatagcttctataagcaggcaactgaaggaccctgtaaactttca
aggcctggattttgggatcctattggaagatataaatgggatgcttggagttcactgggtgatatg
accaaagaggaagccatgattgcatatgttgaagaaatgaaaaagattattgaaactatgccaatg
actgagaaagttgaagaattgctgcgtgtcataggtccattttatgaaattgtcgaggacaaaaag
agtggcaggagttctgatataacctcagtccgactggagaaaatctctaaatgtttagaagatctt
ggtaatgttctcacttctactccaaacgccaaaaccgttaatggtaaagctgaaagcagtgacagt
ggagccgagtctgaggaagaagaggcccaagaagaagtgaaaggagcagaacaaagtgataatgat
aagaaaatgatgaagaagtcagcagaccataagaatttggaagtcattgtcactaatggctatgat
aaagatggctttgttcaggatatacagaatgacattcatgccagttcttccctgaatggcagaagc
actgaagaagtaaagcccattgatgaaaacttggggcaaactggaaaatctgctgtttgcattcac
caagatataaatgatgatcatgttgaagatgttacaggaattcagcatttgacaagcgattcagac
agtgaagtttactgtgattctatggaacaatttggacaagaagagtctttagacagctttacgtcc
aacaatggaccatttcagtattacttgggtggtcattccagtcaacccatggaaaattctggattt
cgtgaagatattcaagtacctcctggaaatggcaacattgggaatatgcaggtggttgcagttgaa
ggaaaaggtgaagtcaagcatggaggagaagatggcaggaataacagcggagcaccacaccgggag
aagcgaggcggagaaactgacgaattctctaatgttagaagaggaagaggacataggatgcaacac

ttgagcgaaggaaccaagggccggcaggtgggaagtggaggtgatggggagcgctggggctccgac
agagggtcccgaggcagcctcaatgagcagatcgccctcgtgctgatgagactgcaggaggacatg
cagaatgtccttcagagactgcagaaactggaaacgctgactgctttgcaggcaaaatcatcaaca
tcaacattgcagactgctcctcagcccacctcacagagaccatcttggtggcccttcgagatgtct
cctggtgtgctaacgtttgccatcatatggccttttattgcacagtggttggtgtatttatactat
caaagaaggagaagaaaactgaactgaggaaaatggtgttttcctcaagaagactactggaactgg
atgacctcagaatgaactggattgtggtgttcacaagaaaatcttagtttgtgatgattacattgc
tttttgttgtccagtagtttagtttgtgtacatatatacacatatatattttgcactacacaaacg
ataacattttaaggactaatattgctgatacttgaataatcaatctctactaggttataagtagta
tacacagatttaccctgcccttgaacttgaaggacattaaattattaatgatcatttggtaacatg
tttacctgattatcttccatagagtaacataagctgcttttcaaaggtaccattgtgataatgaga
tcaaatttataagttattatttttaattttctaaattaaataaaagaaagaatgcaaaaaaaaaaa
aaaaaaaaaa

Seq ID 23

MSVDEKPDSPMYVYESTVHCTNILLGLNDQRKKDILCDVTLIVERKEFRAHRAVLAACSEYFWQAL
VGQTKNDLVVSLPEEVTARGFGPLLQFAYTAKLLLSRENIREVIRCAEFLRMHNLEDSCFSFLQTQ
LLNSEDGLFVCRKDAACQRPHEDCENSAGEEEDEEEETMDSETAKMACPRDQMLPEPISFEAAAIP
VAEKEEALLPEPDVPTDTKESSEKDALTQYPRYKKYQLACTKNVYNASSHSTSGFASTFREDNSSN
SLKPGLARGQIKSEPPSEENEEESITLCLSGDEPDAKDRAGDVEMDRKQPSPAPTPTAPAGAACLE
RSRSVASPSCLRSLFSITKSVELSGLPSTSQQHFARSPACPFDKGITQGDLKTDYTPFTGNYGQPH
VGQKEVSNFTMGSPLRGPGLEALCKQEGELDRRSVIFSSSACDQVSTSVHSYSGVSSLDKDLSEPV
PKGLWVGAGQSLPSSQAYSHGGLMADHLPGRMRPNTSCPVPIKVCPRSPPLETRTRTSSSCSSYSY
AEDGSGGSPCSLPLCEFSSSPCSQGARFLATEHQEPGLMGDGMYNQVRPQIKCEQSYGTNSSDESG
SFSEADSESCPVQDRGQEVKLPFPVDQITDLPRNDFQMMIKMHKLTSEQLEFIHDVRRRSKNRIAA
QRCRKRKLDCIQNLECEIRKLVCEKEKLLSERNQLKACMGELLDNFSCLSQEVCRDIQSPEQIQAL
HRYCPVLRPMDLPTASSINPAPLGAEQNIAASQCAVGENVPCCLEPGAAPPGPPWAPSNTSENCTS
GRRLEGTDPGTFSERGPPLEPRSQTVTVDFCQEMTDKCTTDEQPRKDYT

Seq ID 24

GCGGCCGCCGCCCTCGCCACCCCGCCTGCCCACTCCCGCCGCCGCCCCGCTCTCGCTTTCCCCCCG
GCCTCCCCTCGCCCCTTCCCCTCCCCCTTCCCGGCGCACTCGGGGGGCTGGGAACGAGCTGCCATG
TGATGCGCGTCCCCTCCGCGAGCTTTCGGTGACCCACGAACTGCCCACCTCGCCGGCTGCCGGGAG
GGGGCTGCGAGCCGGGAAGACGCGGGGAAGAGGAGGCGGAAAAGGACGCAAAGTTCTCCGGCGAGC
GCATTCATTCACATAGCTCCCAGTTTTAACATTTCGCCACCTACTGAAGACATCATTTGGGACCAA
GCTGATGAGCCCTTGAAGCACAGCAGATAAGAGTGTTGCTGTTGATCATCTTTGCCTGGGAAGTTG
AATGATAAAGCCAGAAGAAAGCATGCTTTCTGATCATTTGCAGCTGTCTGCTTCAGAAAGTGAGGG
CTCCAGGAATGAGGAGAATCTTCAAGAACTTCCTCGCACTGTGACATGTCTGATCCCTTGCTCCCA
TCCCTGCAGCATGAACAAGGTGGACACTCACTGACCTGTCACAAGGTTGCCCCACAAAACTTTGGG
GTCCATGTCTGAATGGATTGCCAGAGCCTTCTCATCTCTCCCTTCGCCCAGTTCCCTGCATCCTAA
GACTCGAAGGCAGCACAGGACCTGGAAAAATTACATGGTGTGAACGGCATGTCTGTGGATGAGAAG
CCTGACTCCCCCATGTATGTGTATGAGTCCACAGTCCACTGCACCAACATCCTCCTGGGCCTCAAT
GACCAGCGGAAAAAGGATATTCTCTGTGACGTGACTTTGATCGTGGAGAGGAAGGAGTTCCGGGCC
CACCGGGCTGTGCTGGCCGCATGCAGTGAATATTTTTGGCAGGCGCTGGTTGGACAGACAAAAAAT
GATTTGGTGGTCAGCTTGCCTGAGGAGGTCACAGCCAGGGGCTTTGGGCCGCTGTTACAGTTTGCC
TACACTGCCAAGCTGTTACTCAGCAGAGAAAACATCCGCGAGGTCATCCGCTGTGCTGAGTTCCTG
CGCATGCACAACCTGGAGGACTCCTGCTTCAGCTTCCTGCAGACCCAGCTCCTGAACAGTGAGGAT
GGCCTGTTTGTGTGCCGGAAGGATGCTGCGTGCCAGCGCCCACACGAGGACTGCGAGAACTCTGCA
GGAGAGGAGGAGGATGAAGAGGAGGAGACGATGGATTCAGAGACGGCCAAGATGGCTTGCCCCAGG
GACCAGATGCTTCCAGAGCCCATCAGCTTTGAGGCCGCCGCCATCCCCGTAGCAGAGAAGGAAGAA
GCCCTGCTGCCCGAGCCTGACGTGCCCACAGACACCAAGGAGAGCTCAGAAAAGGACGCGTTAACG
CAGTACCCCAGATACAAGAAATACCAGCTTGCATGTACCAAGAATGTCTATAATGCATCATCACAC
AGTACCTCAGGTTTTGCAAGCACATTCCGGGAAGATAACTCTAGCAACAGCCTCAAGCCGGGGCTT
GCCAGGGGGCAGATTAAAAGTGAGCCGCCCAGTGAAGAGAATGAGGAAGAGAGCATCACGCTCTGC
CTGTCTGGAGATGAGCCTGACGCCAAGGACAGAGCGGGGGATGTCGAGATGGACCGGAAACAGCCC

AGCCCTGCCCCTACCCCCACGGCCCCAGCTGGGGCCGCCTGCCTGGAGAGATCCAGGAGCGTGGCC
TCGCCCTCCTGCTTAAGGTCTCTGTTCAGCATAACGAAAAGTGTGGAGCTGTCTGGCCTGCCCAGT
ACATCTCAGCAGCACTTTGCCAGGAGTCCAGCCTGCCCTTTTGACAAGGGGATCACTCAGGGTGAC
CTTAAAACTGACTACACCCCTTTCACAGGGAATTATGGACAGCCCCACGTGGGCCAGAAGGAGGTG
TCCAACTTCACCATGGGGTCGCCCCTCAGGGGGCCTGGGTTGGAGGCTCTCTGTAAACAGGAGGGA
GAGCTGGACCGGAGGAGCGTGATCTTCTCCTCCAGCGCTTGTGACCAAGTGAGCACCTCGGTGCAT
TCTTATTCTGGGGTGAGCAGTTTGGACAAAGACCTCTCTGAGCCGGTGCCAAAGGGTCTGTGGGTG
GGAGCCGGCCAGTCCCTCCCCAGCTCGCAGGCCTACTCCCACGGTGGGCTGATGGCCGACCACTTG
CCAGGAAGGATGCGGCCCAACACCAGCTGCCCGGTGCCAATCAAAGTCTGCCCTCGCTCACCCCCC
TTGGAGACCAGGACCAGGACTTCCAGCTCCTGCTCTTCCTATTCCTACGCGGAGGACGGGAGCGGG
GGCTCACCCTGCAGCCTCCCTCTCTGTGAGTTCTCCTCCTCGCCCTGTTCCCAGGGAGCCAGATTC
CTTGCCACAGAACATCAGGAACCAGGCCTGATGGGAGATGGAATGTACAACCAAGTGCGGCCCCAA
ATTAAATGTGAGCAGTCTTATGGAACCAACTCCAGTGACGAATCCGGATCGTTCTCGGAAGCAGAC
AGTGAGTCGTGTCCTGTGCAGGACAGGGGCCAGGAGGTAAAACTTCCTTTTCCTGTAGATCAAATC
ACAGATCTTCCAAGGAACGATTTCCAGATGATGATTAAAATGCACAAGCTAACCTCAGAACAGTTA
GAGTTTATTCATGATGTCCGACGGCGCAGCAAGAACCGCATCGCGGCCCAGCGCTGCCGCAAAAGG
AAACTGGACTGTATTCAGAATTTAGAATGTGAAATCCGCAAATTGGTGTGTGAGAAAGAGAAACTG
TTGTCAGAGAGGAATCAACTGAAAGCATGCATGGGGGAACTGTTGGACAACTTCTCCTGCCTTTCC
CAGGAAGTTTGCCGAGACATCCAGAGCCCCGAGCAGATCCAGGCCCTGCATCGGTATTGCCCTGTC
CTCAGACCCATGGACTTGCCCACGGCCTCCAGTATTAACCCTGCGCCCTTGGGTGCTGAGCAGAAC
ATTGCGGCCTCCCAATGCGCAGTGGGGGAAAACGTGCCCTGCTGCTTGGAGCCAGGCGCGGCTCCC
CCCGGACCCCCCTGGGCACCCAGCAACACCTCCGAGAATTGTACCTCTGGGAGGAGACTAGAAGGC
ACTGACCCGGGAACCTTCTCAGAGAGAGGACCTCCTCTTGAACCCAGGAGCCAAACAGTGACCGTG
GACTTCTGCCAGGAAATGACTGATAAGTGTACAACTGACGAACAGCCCAGGAAAGATTATACCTAG
TGACTCGGCTCTGCCTCCCAGTCCGCACACCTCTCCCATCCAGGCGTTCTTCAGTCAGCCTGTGGC
ACTGTTCATCTGCTGTCCCGAAGAAACCGAGAACACATTTGGTGCACACTACAGCGGTCTTAGCAG
CAATACTGTTCCGAAGTATCCTCTCCTCTTCTCGAGCAGGAGTGATAGTTACCTTCACAATGGTGC
TACCCCTTGCCCAGGCAAGGAAAGACAGCAGTGATGACACTGTCTGTCTGTGGCTCAATTTCAGTC
TTCACAGGGATAGACTACAACACCTCTAGGCCCCAACCACGGATTTTTTTTCTCAGTGGCCCATGT
CACAAACCCTATCTCAGGAATTTCTTCTGAATGTTCAATTTTTTTCATTGAAGACAGCTTCTATAC
ACATCAAAGTTTTATAGCTAGACTGTACATATTATATATAATATATATATAAAAAATATATATATA
TATATATATCCATATGCAAAAGTCCTGCATGCCTCAACTTTCTCATCCTAAAACTGGAAACTTATT
TCTCATTTAGAAACAGGTTCCAACATTCCTCTTCTTTTGTCTCTGATGCTAGAACTAGTTTGGTAA
CTGTTAACGTGGTCATTTTTCTTGCTTCACAGTTCAATTTTCAATTCGTACTTATTTATGGACAAA
ATTCAGTGTTGGAAGCTTTTTCCCAAGGTTTTATTTCAGATTTCTTTTTCGTTTGGTTTGGTTTTG
GCACCTCCAAGTGGTGTCATTTGAGCATTGTAGGTTTGTTTTTTGTTTGTTTGGGGGGTTTTGTTT
GTTTTTGTTTTTGTTTTTGTTTTCCTTGCAGATACTGTACAGTAATGGTCAACTTTGCCACTTGCA
CTGAGTTTTGGGTCAAACCTATTTTCTTAAATGAAGTTGTAACTTCGGTATAACTCAAGTATACTG
TATATTCTTTGCTTTTAGTTAAAAAAGTAAAACATTTTAGCTAATTAAAAAGCACTCAGGTGATAA
TTATGTAGGAAAAACAATCTTGCCAAATAATGAATTCATCCTAGGATGTGTAGACAATAATCTGCT
TGAATATTTTTATATTTCACCTCCTCCCCACCTTTCCCTAAGCAAAGTTTAAACGCAGATAGAGAG
TTCAGAGTTGATGCTGGATGTTCAGATTCCTAAGTGGGGAGAGAGTTTGGACATCTCACTCAAAAG
TACATCAGAAAAACAGGAATCCGTGATTTTATACCAGAACTCAGCAGGCATTGGCTCCTAGAAATC
AAGTTAGAAAGTTTTCACCCAGGGAGTAAGTCCCATTCATTTCAACACGTCCTGAGGCCTCGGCTT
GCTCTTGGAAGTGGTGTGCAGTAGGACCTGCTCCCCTGAAGGACGGGGCCAACCAGCCACTGGCTT
TCCTGCCCAGGCTTGGCCTCCCAGGACATCTGGCCTGAGGGGATTTGAATCACAGCCCCGAAGGTC
CTGCCTTCACCCCATTGGGAGAGAGCAGGGCATCCTGGCATCTGCGATCCATCCCTGACACAGGCT
GACACATTCTTTCTCCTTTCCTTCTCCAAAGGCTTGGAGTTTTCTTCTGAGGTTTTTCTGCCAGTG
TCTTGTCTGAAGGCAGACTTCATTCTGAGGCTTTGGACAAGCTATTCACCGGGAACCCTCCCTGTC
CCCTTCCCGAATCACACACATACCCTACCCTCACCTGATGATAATTTTCTCTTCTTGCTGCAAAAC
TGGTTGGCTTGCAACCCAGAGAGAGCAGCTTCCCTTGGCTCTGGGGGCCGTGTTGGCCCCAGCCAC
GTTTACAGGAAGGTGTGCCCCAGAGGAGGAGGAATCAGCTCCCTCGCTCCAGTGGCCTTGGGTCCG
GGTCTCACTGAGCAGCCCGAGGGCCACTCCAGCCCGGCTGGGGAAGAGAGTCCTGAACGGTTTGAT
GTGGGGATGGGGTGGTGGGCAGTGGGGAATAGATGGTTGACTTTGTTTCTTTATTTGTGCCATTGT
TTGGACAATATTAAAGCTGCATGTAAAAGGGGAAATTAGTATATGATGTAGGCTAAAAGTGAAATC

ATAGTAACATATGTTTTAGTATTATTAACTTTTTTCTGTACAAATATTAGCACTAAATGTTTAAAT
ATGTATGAATGCCAGAAATTTGTCAGTTCATGCAGTAGGATAAAAAAAAAAAAAAAAAAAAGGG
CTTTTCTTTTTAAACAGTTCCACTTTTAAAACCTGCCTCTGGGTTTTTGTTTTTTCTTGTTTGTGT
GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCTGAAACAGATCTTGATAAAGC
TCTGTGTTGGAGCTGCTGGTTTTTGTTATGGTTGTTGGAATTTCTTGGCCTACTAGGACAGTTCTG
TGCTTCACCATGAGGTTTGCCTTTGTGGAAAACTCGTGGGTGACAGTGAGAATATAAACTCAATGT
GAATCACGTGATACTTCGGCAGGCGTGTGTTACAGTGGAGTCAGCTGACAGTATTTTGCTTTTTAA
CTCTATTGTTGCCTTTCCAAGTGACCTCTCCTCTTCTTTTAAAAAAAGAACACTTTCTGCTCATAT
CATAACCAGGTCCAACCCAGCTTCTTGGCATGAGGTTTACCCTGGTAACAACTCATGTGCAACTGG
TAGTCTTGACCACATTCCATCCATTTCCTCAGGTTTCTGTGGTTCAGTAGCCCAGACCTGTTTGGC
AGCCATTTCTAGCAGGGGCGGGGCCTCTTTATTTCTCTCCACCCTAACTCAGACCTCACCTTCCTC
CCACCCACCCCTGCCTTGCTTTTCTTCCTCTTCCCCCAACCTAACTTCTGCCATGGGAACTGGTTA
AAAACACTGCTCTAAAAACCATCTTCCAATTTCATAGAGATTTCTCACAAGTTATTTCATTCATAA
TCCACCATGAACAGTGACTAGCTTCGTGCAGTTGTTCATGTGATGTGTGTGTCTTTTCCTATTC
AGAACTATGTGCTTGTCAAAATTATTTCTGGGTTGATTCAAAGGGAGGACTTGCTGGGGACCAGAA
TCCAAACGGCCTCAAGTGGAATTTTAAAACCTAGCCTGTCTCTTTTCCCTGGGATCCCTCTGTCAA
CCCCACGCCTTTTAGGAAAAAGAAAAGTGAGTGAACAGCAAGGAAGAGTGTTTGCACAGTACAGTA
ACATTTGGTTGTTCTTAAGGCTCTTTTCTTACAAAAATAAGAGACCCTCCAACCACGGGCTGTTTA
GGAGGATGCCTGCTTGGGTCTCCAAATGGCTGGGGTAGGAATGGTTGTTGGGGCAGAGCCAGTGGA
GGTGAGTGACCCTGAGACTAATGAACATCCCACCTAAATCCAGTCCTCCCCTTGGATCTGCCTTTG
TCCTGCTTGTGTATCCAGGCAACCTCTTTTCAAGTTGGTCAGGCTTTGGACAGGTGAGTGATTTGC
TGTATGTGTTTGTTTCTCTGCGTTACCTGGGGGGTGCCTTGATTAAAATCGAACTTTATTACATACT
GATTCTGGAACAAAACAGTTAGAAAAACTTTAAACTTTAAAAAAAAAAAAACCGACAAAGTTACGA
GGCCATCCTGCTATTTATCTTCTGAGTTCCCAGCAATGACTCAGGCATCAGAGATGATGCTGCAGT
GGAAAACCTGACTCTGTGTGTCTGCAACTGAATGTTGTGCGAGTAATTTATTAACTGTCTTTCTAA
AGGTTTGCTGCTTTTAAGATGCACTATAATTCGGGATGTAATCCTTACATTGCTTTTCCAAGGAAG
GGAACAAAAGTCTAGTGATTAGTATGCCAACTGCCACTACTCCTTCAAAAGGAGCCAGGACCAGCG
ACAAGACTCATGAGAGGACTGGCTAAAGTGAAGTGTGCACAGTGTGAAGTTTAATGCTGTTGTCAA
GAGGCCTAAACCCACATTTTCTCTTTTAATATTTTATGATTGCCATCAAAGAAGAAGAAAAAGAAG
GAACAGACAAAGGTTTGAAAATGATAAGCCTGTTAAGACACCAAAAACTCCTGTCCCGTGAAGCTG
CTTGACATCCTGTGGAGTAGCATAATCCTCTCAAAATGAGGAAGAGCTGCCTGCAAAGCTTTCTCA
AGTCCCTATTTGGCTACCTACTTCTCTACATTATGCCCCATTTAAACTAGGAGCTGTCTTAGAAAT
GACTTCAAACTGCTTCACTATTGCTTACAGTTTAGGAGGAGTCTCAGATCCAGAAGGAGCAAGAAT
CAAGTTTGGTCCTCAAATGACTGTAAATAGACTAAAGAACAAGGTGTTTTTTGTTTTTTGTTTTTGT
TTTCTAAGAATAAAGCTGTTCGTTGTATCATGAGTTAGTGTTTCTTCCCCAAACTGAAGACTGTGT
TGGAAGTGCAATTTCTGGTGAGTCAGTCCACAATACAATGCCCTGTGTGGAGTTGGTATTCATACA
GGAAATCTGTGTGCACGAGGCATTGTGTGTTGAAAGTGTATGTTTATAGTACTGCCTGAGCCATCT
CATGACCCCAGCGTCCAAAACCGATGCTGTAGAACAGAACATATCTGTCACAAATAGGTGTGTGCA
AATAGCATTTGTACATAGAAAAGTCTCATTGTGGCAGATTGAGCATAAATTATTCAACTGACGGTG
CAAAAACATTACTTGCAAAGAAAAGTTTATAGTATTTTCCTACACTCCACCCTGGGAGATGATATT
TCTATCAAATGAATATCAGTGCATTTTAAATGTAATATGAAAACGATGCTGCCATTTTGTGAAGAA
TACCCACTTGGTTGCAGAGGCCAACTTTCATAGCTTTGATTTAATGTTGTGACACGGTGTATGCAT
TTTGCTGTCAAGCAATGGATAAACAGCTCTGACTTTCATTCTCATTCCAGTTTATTGACCTCAGAT
AAAACACTGGCCCTTCTTAGAAGCAGAAGTGTGCACCAAGACCATTCATTTCAGGTAGACTCACAT
TCAGTGCCAAGTGCTCCCATGGGAATAATCAGACGCATATGTTGCGAAAGAGTGAAGGGACTTGGA
CAAAGAGGGGTTTTCCTACAGATGGATGCTCAGTCTTCTACCAAAACATGTTTGGAGGCAGAACTA
TGACCTCCCCTTAAGTCCTAACAATGTATTTTGTGTGTGCAAATCCTGGGATGCCCGTTTCACGCT
CTGACATAAAGACATGGCACCTCTAGTGAGTGATCAGGAAGATTCCATATGCATTTGGGAGCTTCA
GGTGCTTGTTAGACACAGTGAGCCATTCAAGGCAAGCACCACCTTTGCTAGTGAGGCCAAGAGAGC
CTGTGACAATTTGACAATTTGTTCCAGAACCAGTCTGATGCAAGTGCACCTCTAATATATGCCTTA
CAAACTCCAGAGGCCATATTCAAAACAGGGTCTTCTCAGTGTATGCAAGGGGCTGCAGCCCCTCTT
CTCTTCCTCCCCAGGTTGAACAATACGGACAGTTTTCACACATATCTACCTGTATAACCCTCTGTA
CCTCTCATAACTGGTCAACGACTGTAACAGGTTACATCAGGTGTTTTTCTACATACTTTTTACACA
GATTCTATGCGATTAATGTAATTTAATTCAATGCATCATTTTATTGTACTAGTTCTTAGGCTTGTC
CTTATTTTTTTCTAAGTGATTGTGGTTTTTCTCGTGGTTTTTATTGTAAAAAATGAAAGGCTGTTG

ATGCTTATTCTCTGTAACTAAGAATTTTACCTTTTGGGGGAAAAAAGCATTGCTATGAACTAATGG
AATTGGAACTTCATTTACTCATTGTAAATACACTATTGTGCAAAAAAAGTTTTCACTCAATTGAAT
TGCTAGTGTTAACTGAATTTTGTCTAGACACCATTTCTGTTGATGAAATAAAGACATATCATTATG
CATTGTAAACTG

Seq ID 26

TTCTAAAGCTACTGGAGCACGGGACACGTGTCTGTACGCTCCTCTCACACCCCTTACTCGGCAGCT
GCCCACTCCCGCCGTTGCCGTTGCCCTGAGTCTTCCCCACTGTTCTGCTGTGCCAGTGTGAGCATC
TCCCCGGCACTCAAGTCCATGTTCTCTCCCTCCCGCTCTTCCTCCTGTCTCTTCCATGCTCCCTGT
TGCACGGGGACATCATGTGGGGCTTGATGCCGTACACAGTCCCTGGACCATGGGCCCCCTAACTTC
TTTCTTCCAGGAAGTGTTCACACTAGGACAGGGAGGTGGTTGGAGGTCACCAACTTCACCTTTTCC
CTAGGAATGTGGGGGTTTTGTTTTTTTCTTCCCCCCATTCATTTGGCAGGAGGGTTTTTTTTTTTC

Seq ID 27

MAGTNALLMLENFIDGKFLPCSSYIDSYDPSTGEVYCRVPNSGKDEIEAAVKAAREAFPSWSSRSP
QERSRVLNQVADLLEQSLEEFAQAESKDQGKTLALARTMDIPRSVQNFRFFASSSLHHTSECTQMD
HLGCMHYTVRAPVGVAGLISPWNLPLYLLTWKIAPAMAAGNTVIAKPSELTSVTAWMLCKLLDKAG
VPPGVVNIVFGTGPRVGEALVSHPEVPLISFTGSQPTAERITQLSAPHCKKLSLELGGKNPAIIFE
DANLDECIPATVRSSFANQGEICLCTSRIFVQKSIYSEFLKRFVEATRKWKVGIPSDPLVSIGALI
SKAHLEKVRSYVKRALAEGAQIWCGEGVDKLSLPARNQAGYFMLPTVITDIKDESCCMTEEIFGPV
TCVVPFDSEEEVIERANNVKYGLAATVWSSNVGRVHRVAKKLQSGLVWTNCWLIRELNLPFGGMKS
SGIGREGAKDSYDFFTEIKTITVKH

Seq ID 28

AATTCTAGGGACAACCTGAGTGCTCAGTCGTAAAGAGGAAAGGCAGAATTTTTCCTTGCTATGGCT
GGAACAAACGCACTTTTGATGCTGGAAAACTTCATAGATGGAAAATTTTTACCTTGTAGCTCATAT
ATAGATTCTTACGACCCATCAACAGGGGAAGTGTATTGCAGAGTGCCAAATAGTGGAAAAGACGAG
ATCGAAGCCGCGGTCAAGGCCGCCAGAGAAGCCTTTCCCAGCTGGTCATCCCGCAGCCCCCAGGAG
CGCTCACGGGTCCTGAACCAGGTGGCGGATTTGCTGGAGCAGTCCCTGGAGGAGTTTGCCCAGGCC
GAGTCTAAAGACCAAGGGAAAACCTTAGCACTGGCAAGAACCATGGACATTCCCCGGTCTGTGCAG
AACTTCAGGTTCTTCGCTTCCTCCAGCCTGCACCACACGTCAGAGTGCACGCAGATGGACCACCTG
GGCTGCATGCACTACACGGTGCGGGCCCCGGTGGGAGTCGCTGGTCTGATCAGCCCCTGGAATTTG
CCACTCTACTTGCTGACCTGGAAGATAGCTCCAGCGATGGCTGCAGGGAACACTGTGATAGCCAAG
CCCAGTGAGCTGACTTCAGTGACTGCGTGGATGTTGTGCAAACTCCTGGATAAAGCAGGTGTTCCA
CCAGGTGTGGTCAATATTGTGTTTGGAACCGGGCCCAGGGTGGGTGAGGCCCTGGTGTCCCACCCA
GAGGTGCCCCTGATCTCCTTCACCGGGAGCCAGCCCACCGCTGAGCGGATCACCCAGCTGAGCGCT
CCCCACTGCAAAAAGCTCTCCCTGGAGCTGGGGGGCAAGAATCCTGCCATCATCTTTGAGGACGCC
AACCTGGATGAGTGCATTCCGGCAACCGTCAGGTCCAGCTTTGCCAACCAGGGTGAAATCTGTCTC
TGTACCAGCAGGATCTTTGTCCAGAAGAGCATCTATAGTGAATTTTTAAAGAGATTTGTAGAAGCT
ACCAGAAAGTGGAAAGTCGGCATTCCCTCTGATCCACTGGTGAGCATAGGTGCTCTGATAAGTAAA
GCACATTTGGAGAAAGTCAGAAGTTACGTCAAGAGAGCTCTTGCTGAAGGTGCCCAAATTTGGTGC
GGTGAGGGAGTGGATAAGTTGAGCCTCCCTGCCAGGAACCAGGCAGGCTACTTTATGCTTCCCACG
GTGATAACAGACATTAAGGATGAATCCTGCTGCATGACGGAAGAGATATTTGGTCCAGTGACGTGT
GTCGTCCCCTTTGATAGTGAAGAGGAGGTGATTGAAAGAGCCAACAACGTTAAGTATGGGCTGGCG
GCTACCGTGTGGTCCAGCAATGTGGGGCGCGTCCACCGGGTGGCTAAGAAGCTGCAGTCTGGCTTG
GTCTGGACCAACTGCTGGCTCATCAGGGAGCTGAACCTTCCTTTCGGGGGGATGAAGAGTTCTGGA
ATAGGTAGAGAGGGAGCCAAGGACTCTTACGACTTCTTCACTGAGATCAAAACCATCACCGTTAAA
CACTGATCTTTGCTAATGGTGGAGCCACTATGGCCAATGCCTGGCTGCAGGCATCAATTGTTCAAT
GTGGTAGATGAAAATCATGGCATGAATTCCAGCTATGCCTTGACCTGGCAGAAGGTTATCTCTAGC
TTATCCTCAGTTCTTAGTAACTTTACCCACTAGTGAAGAGATACTGTCTATTTTCAATGTGGACTC
GGAAAAAAAGACTTATAAGTAGGAAGACAGAACAATGATGCCAGTTGTCAGGCTCCTCCCAGGTTA
TGTTTTCATAGTGTTTCTTTCATCATCTTCATTGAACTCTTGGGAATCTCCAGATAATCAGATTAT
TTCATTTGGTAAATTTTAAAAAATATGCAATCAGGCACAGTGCCTCATGCCTATAATCCCAGCACT
TTGGGAGGCCAAGGTGGGTGGATCACTTGAGTTCAGGAGTTCGAGATCAGCCTAGGCAACATGGTG
AAATCCTGTCTTTACCAAAAGTTTAAAAATTAGCTTGGTGTGGTGCCCTCTGCCTATAGCCCCAGC

TACTTGGGAGGCTGAGGTGGGAGGATCGCTTGAGCCCAGGCGGTTGAGGCTGCAGTGAGCCATGAT
CATTCCACTGCATTTCAGCCTGGGGGATACAGTGAGACCTTGTCTTTAAAAAAAAAATGCTGCAAA
ACCAAAAATAAATAGCCATAATTTGCCATCTTTGGAGAGATTATGGAGATAATCTCCTTGTCTATT
AGCCACTAGCATGGCTAATTGATCATTGATTGGACCTCATTACATAAGGGGGCTGTGGACGTCTTT
GGAATTTGTTAGGGGAAGGCTATGTGGTAAGCTCTACTAAATGTATCTGTACAAGGATGAAGAAAA
ACAAAATAACTACTACTTTTGGAGGAAACTTGGAACAAGAAGAGTACATAGTTTGATCCATTCCTA
CGGTCGAGTCTAATCAATTCCATTTCTTAGTGTATGCATCTAATGAGGCTTACAACTTGCTAACCC
CAAAAGAAGCAATAAATGTTAAGCAATTTTAATAAAAAAAAAA

Seq ID 30

TCCAGGGTATACGTGCAGGATGTGCAGGTTTGTTACACAGGTAAACATGTGCTATGGTGGTTTGCT
GTACCACACCCTGGTAATCTTTATTTTTTGTAGAGATGAGGTCTCACTATGTTGCACAGAGGGATC
TCGAACTCCTGGGCTCAAGGGCTCACAGCCTCCCAAAGCACTGGGATTACCCCTGTGAGCAATCAT
GCCTGGCCTTATTTTTGTACAATGTTGCTAGTAATCCCTACCTCAGCAGGCTGTGGTAAGGNTTAA
AAGGTACATTGTACATAAAGTACATTTTCTAGTACGTTTAATAAATGTTCATTTCTATCTTCCTGA
AAATTTCTCAGAGGAANTGCGAAGGAGTGGGACACT

Seq ID 32

GAGACTGCATAGGGCTCGGCGTGGTATTGATGCAAGGATAGGTATATATATCAATGGAAGAGAATC
GAGAGTTCACAAATAAACTCTTACAATTATAGCCAATGATTTTCAACATAGGTGTCCAAACAATTC
AATGGAAGAAAGGAGAGTCTCTTTTTTTCACATTCAGAATTTTTATTTTAAAACAAAGAATCGG

Seq ID 34

GACNCTGCATAGGGTNNGGNGTGGGGAACAAATCTGGTCAGTTGGCGTTAGTTTGCTACTTTACAG
CCTCCTGGATTTAAAATTCAAAGGTATTATTGGGAGAACAGTACAAATTTGAAAAGACCTGTAGAT
TAAATAAGAGTATTGTTTCAAGGATAATTTTTTGATTGTTAGGGATAAGTGGGCATTGTGTTTACA
AATTACTTCCAAAGAATTCAGAAAATTGTGTGTGTGGGAGGCAGGGTAGCAAGATAAAAAGAGGGA
GGACAGCTGGGGTTGGTAAAAACAAATGTAAAATGACAACATTTGGGAGAAACCTGGGTGTAATGA
TTACTTTTGTGTCCAATTTGGCTAGGGCACAATACCANTATCCTATGTACTTTAAGCCAGATACTT
TAAGTCGAAGGGAGTAACCANTTAATTCCAATTTAAACACTTATTCCCGGGTGTTCAGCCTGAAGG
GTATTTTATACCATGTGGGTTANCATCCTATAGGNCAGCCNCACCTTTTTTAAGGGGATTTA

Seq ID 37

MDPNCSCSTGGSCTCSSSCGCKNCKCTSCKKSCCSCCPVGCSKCAQGCVCKGASDKCTCCA

Seq ID 38

TCCCGACTTCAGCAGCCCGACTGCCTTCTTGTCGCTTACACCGTTGCTCCAGATTCACCAGATCTC
GGAATGGACCCCAACTGCTCCTGCTCCACCGGCGGCTCCTGCACCTGCTCCAGCTCCTGCGGCTGC
AAGAACTGCAAATGCACCTCCTGCAAGAAGAGCTGCTGCTCCTGCTGCCCCGTGGGCTGCTCCAAA
TGTGCCCAGGGCTGTGTCTGCAAAGGTGCCTCGGACAAGTGCACGTGCTGTGCCTGAAGTGACGAA
CAGTGCTGCTGCCCTCAGGTGTAAATAATTTCCGGACCAACTCAGAGTCTTGCCGTACACCTCCAC
CCAGTTTACTAAACCCCGTTTTCTACCGAGCATGTGAATAATAAAAGCCTGTTTATTCT

SEQ ID NO:39

Nucleotide sequence of ires-GFP DNA fragment

CTAGAGTGTGATTTTAAGGGCGAATTCTGCAGATATCCATCACACTGGCGGCCGCACTAGAGGAAT
TCGCCCCTCTCCCTCCCCCCCCCCTAACGTTACTGGCCGAAGCCGCTTGGAATAAGGCCGGTGTGT
GTTTGTCTATATGTGATTTTCCACCATATTGCCGTCTTTTGGCAATGTGAGGGCCCGGAAACCTGG
CCCTGTCTTCTTGACGAGCATTCCTAGGGGTCTTTCCCCTCTCGCCAAAGGAATGCAAGGTCTGTT
GAATGTCGTGAAGGAAGCAGTTCCTCTGGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCT
TTGCAGGCAGCGGAACCCCCCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGA
TACACCTGCAAAGGCGGCACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAA
ATGGCTCTCCTCAAGCGTAGTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCCATTGTATGGG

AATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAAGCTCTAGG
CCCCCCGAACCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATACCATGGTGAGCAAGGGCG
AGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGT
TCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCA
CCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCT
TCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACG
TCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCG
AGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCC
TGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGA
ACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACC
ACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCA
CCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGA
CCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGCGACT

SEQ ID NO:40
Nucleotide sequence of DNA fragment containing human HIF-1α protein coding sequence

CTAGCCGTAGAATCCGACCGATTCACCATGGAGGGCGCCGGCGGCGCGAACGACAAGAAAAAGATA
AGTTCTGAACGTCGAAAAGAAAAGTCTCGAGATGCAGCCAGATCTCGGCGAAGTAAAGAATCTGAA
GTTTTTTATGAGCTTGCTCATCAGTTGCCACTTCCACATAATGTGAGTTCGCATCTTGATAAGGCC
TCTGTGATGAGGCTTACCATCAGCTATTTGCGTGTGAGGAAACTTCTGGATGCTGGTGATTTGGAT
ATTGAAGATGACATGAAAGCACAGATGAATTGCTTTTATTTGAAAGCCTTGGATGGTTTTGTTATG
GTTCTCACAGATGATGGTGACATGATTTACATTTCTGATAATGTGAACAAATACATGGGATTAACT
CAGTTTGAACTAACTGGACACAGTGTGTTTGATTTTACTCATCCATGTGACCATGAGGAAATGAGA
GAAATGCTTACACACAGAAATGGCCTTGTGAAAAAGGGTAAAGAACAAAACACACAGCGAAGCTTT
TTTCTCAGAATGAAGTGTACCCTAACTAGCCGAGGAAGAACTATGAACATAAAGTCTGCAACATGG
AAGGTATTGCACTGCACAGGCCACATTCACGTATATGATACCAACAGTAACCAACCTCAGTGTGGG
TATAAGAAACCACCTATGACCTGCTTGGTGCTGATTTGTGAACCCATTCCTCACCCATCAAATATT
GAAATTCCTTTAGATAGCAAGACTTTCCTCAGTCGACACAGCCTGGATATGAAATTTTCTTATTGT
GATGAAAGAATTACCGAATTGATGGGATATGAGCCAGAAGAACTTTTAGGCCGCTCAATTTATGAA
TATTATCATGCTTTGGACTCTGATCATCTGACCAAAACTCATCATGATATGTTTACTAAAGGACAA
GTCACCACAGGACAGTACAGGATGCTTGCCAAAAGAGGTGGATATGTCTGGGTTGAAACTCAAGCA
ACTGTCATATATAACACCAAGAATTCTCAACCACAGTGCATTGTATGTGTGAATTACGTTGTGAGT
GGTATTATTCAGCACGACTTGATTTTCTCCCTTCAACAAACAGAATGTGTCCTTAAACCGGTTGAA
TCTTCAGATATGAAAATGACTCAGCTATTCACCAAAGTTGAATCAGAAGATACAAGTAGCCTCTTT
GACAAACTTAAGAAGGAACCTGATGCTTTAACTTTGCTGGCCCCAGCCGCTGGAGACACAATCATA
TCTTTAGATTTTGGCAGCAACGACACAGAAACTGATGACCAGCAACTTGAGGAAGTACCATTATAT
AATGATGTAATGCTCCCCTCACCCAACGAAAAATTACAGAATATAAATTTGGCAATGTCTCCATTA
CCCACCGCTGAAACGCCAAAGCCACTTCGAAGTAGTGCTGACCCTGCACTCAATCAAGAAGTTGCA
TTAAAATTAGAACCAAATCCAGAGTCACTGGAACTTTCTTTTACCATGCCCCAGATTCAGGATCAG
ACACCTAGTCCTTCCGATGGAAGCACTAGACAAAGTTCACCTGAGCCTAATAGTCCCAGTGAATAT
TGTTTTTATGTGGATAGTGATATGGTCAATGAATTCAAGTTGGAATTGGTAGAAAAACTTTTTGCT
GAAGACACAGAAGCAAAGAACCCATTTTCTACTCAGGACACAGATTTAGACTTGGAGATGTTAGCT
CCCTATATCCCAATGGATGATGACTTCCAGTTACGTTCCTTCGATCAGTTGTCACCATTAGAAAGC
AGTTCCGCAAGCCCTGAAAGCGCAAGTCCTCAAAGCACAGTTACAGTATTCCAGCAGACTCAAATA
CAAGAACCTACTGCTAATGCCACCACTACCACTGCCACCACTGATGAATTAAAAACAGTGACAAAA
GACCGTATGGAAGACATTAAAATATTGATTGCATCTCCATCTCCTACCCACATACATAAAGAAACT
ACTAGTGCCACATCATCACCATATAGAGATACTCAAAGTCGGACAGCCTCACCAAACAGAGCAGGA
AAAGGAGTCATAGAACAGACAGAAAAATCTCATCCAAGAAGCCCTAACGTGTTATCTGTCGCTTTG
AGTCAAAGAACTACAGTTCCTGAGGAAGAACTAAATCCAAAGATACTAGCTTTGCAGAATGCTCAG
AGAAAGCGAAAAATGGAACATGATGGTTCACTTTTTCAAGCAGTAGGAATTGGAACATTATTACAG
CAGCCAGACGATCATGCAGCTACTACATCACTTTCTTGGAAACGTGTAAAAGGATGCAAATCTAGT
GAACAGAATGGAATGGAGCAAAAGACAATTATTTTAATACCCTCTGATTTAGCATGTAGACTGCTG
GGGCAATCAATGGATGAAAGTGGATTACCACAGCTGACCAGTTATGATTGTGAAGTTAATGCTCCT

ATACAAGGCAGCAGAAACCTACTGCAGGGTGAAGAATTACTCAGAGCTTTGGATCAAGTTAACTGA
GCGGATCCGACGGGGATCCT

SEQ ID NO:41

Nucleotide sequence of DNA fragment containing human EPAS1protein coding sequence

AGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCAGCGACAATGACAGCTGACAAGGAGAAGAAA
AGGAGTAGCTCGGAGAGGAGGAAGGAGAAGTCCCGGGATGCTGCGCGGTGCCGGCGGAGCAAGGAG
ACGGAGGTGTTCTATGAGCTGGCCCATGAGCTGCCTCTGCCCCACAGTGTGAGCTCCCATCTGGAC
AAGGCCTCCATCATGCGACTGGAAATCAGCTTCCTGCGAACACACAAGCTCCTCTCCTCAGTTTGC
TCTGAAAACGAGTCCGAAGCCGAAGCTGACCAGCAGATGGACAACTTGTACCTGAAAGCCTTGGAG
GGTTTCATTGCCGTGGTGACCCAAGATGGCGACATGATCTTTCTGTCAGAAAACATCAGCAAGTTC
ATGGGACTTACACAGGTGGAGCTAACAGGACATAGTATCTTTGACTTCACTCATCCCTGCGACCAT
GAGGAGATTCGTGAGAACCTGAGTCTCAAAAATGGCTCTGGTTTTGGGAAAAAAAGCAAAGACATG
TCCACAGAGCGGGACTTCTTCATGAGGATGAAGTGCACGGTCACCAACAGAGGCCGTACTGTCAAC
CTCAAGTCAGCCACCTGGAAGGTCTTGCACTGCACGGGCCAGGTGAAAGTCTACAACAACTGCCCT
CCTCACAATAGTCTGTGTGGCTACAAGGAGCCCCTGCTGTCCTGCCTCATCATCATGTGTGAACCA
ATCCAGCACCCATCCCACATGGACATCCCCCTGGATAGCAAGACCTTCCTGAGCCGCCACAGCATG
GACATGAAGTTCACCTACTGTGATGACAGAATCACAGAACTGATTGGTTACCACCCTGAGGAGCTG
CTTGGCCGCTCAGCCTATGAATTCTACCATGCGCTAGACTCCGAGAACATGACCAAGAGTCACCAG
AACTTGTGCACCAAGGGTCAGGTAGTAAGTGGCCAGTACCGGATGCTCGCAAAGCATGGGGGCTAC
GTGTGGCTGGAGACCCAGGGGACGGTCATCTACAACCCTCGCAACCTGCAGCCCCAGTGCATCATG
TGTGTCAACTACGTCCTGAGTGAGATTGAGAAGAATGACGTGGTGTTCTCCATGGACCAGACTGAA
TCCCTGTTCAAGCCCCACCTGATGGCCATGAACAGCATCTTTGATAGCAGTGGCAAGGGGGCTGTG
TCTGAGAAGAGTAACTTCCTATTCACCAAGCTAAAGGAGGAGCCCGAGGAGCTGGCCCAGCTGGCT
CCCACCCCAGGAGACGCCATCATCTCTCTGGATTTCGGGAATCAGAACTTCGAGGAGTCCTCAGCC
TATGGCAAGGCCATCCTGCCCCCGAGCCAGCCATGGGCCACGGAGTTGAGGAGCCACAGCACCCAG
AGCGAGGCTGGGAGCCTGCCTGCCTTCACCGTGCCCCAGGCAGCTGCCCCGGGCAGCACCACCCCC
AGTGCCACCAGCAGCAGCAGCAGCTGCTCCACGCCCAATAGCCCTGAAGACTATTACACATCTTTG
GATAACGACCTGAAGATTGAAGTGATTGAGAAGCTCTTCGCCATGGACACAGAGGCCAAGGACCAA
TGCAGTACCCAGACGGATTTCAATGAGCTGGACTTGGAGACACTGGCACCCTATATCCCCATGGAC
GGGGAAGACTTCCAGCTAAGCCCCATCTGCCCCGAGGAGCGGCTCTTGGCGGAGAACCCACAGTCC
ACCCCCCAGCACTGCTTCAGTGCCATGACAAACATCTTCCAGCCACTGGCCCCTGTAGCCCCGCAC
AGTCCCTTCCTCCTGGACAAGTTTCAGCAGCAGCTGGAGAGCAAGAAGACAGAGCCCGAGCACCGG
CCCATGTCCTCCATCTTCTTTGATGCCGGAAGCAAAGCATCCCTGCCACCGTGCTGTGGCCAGGCC
AGCACCCCTCTCTCTTCCATGGGGGGCAGATCCAATACCCAGTGGCCCCCAGATCCACCATTACAT
TTTGGGCCCACAAAGTGGGCCGTCGGGGATCAGCGCACAGAGTTCTTGGGAGCAGCGCCGTTGGGG
CCCCCTGTCTCTCCACCCCATGTCTCCACCTTCAAGACAAGGTCTGCAAAGGGTTTTGGGGCTCGA
GGCCCAGACGTGCTGAGTCCGGCCATGGTAGCCCTCTCCAACAAGCTGAAGCTGAAGCGACAGCTG
GAGTATGAAGAGCAAGCCTTCCAGGACCTGAGCGGGGGGGACCCACCTGGTGGCAGCACCTCACAT
TTGATGTGGAAACGGATGAAGAACCTCAGGGGTGGGAGCTGCCCTTTGATGCCGGACAAGCCACTG
AGCGCAAATGTACCCAATGATAAGTTCACCCAAAACCCCATGAGGGGCCTGGGCCATCCCCTGAGA
CATCTGCCGCTGCCACAGCCTCCATCTGCCATCAGTCCCGGGGAGAACAGCAAGAGCAGGTTCCCC
CCACAGTGCTACGCCACCCAGTACCAGGACTACAGCCTGTCGTCAGCCCACAAGGTGTCAGGCATG
GCAAGCCGGCTGCTCGGGCCCTCATTTGAGTCCTACCTGCTGCCCGAACTGACCAGATATGACTGT
GAGGTGAACGTGCCCGTGCTGGGAAGCTCCACGCTCCTGCAAGGAGGGGACCTCCTCAGAGCCCTG
GACCAGGCCACCTGAGCCAGGCCTTCTACCTGGGCAGCACCTCTGCCCACGCCGAGCCCTATGCAG
TCTCGGCCGCAAGCTATCAGATCTGCCGGTCTCCCTATAGTGAGTCGTATTAATTTCGATAAGCCA
GGTT

SEQ ID NO:42
The nucleotide sequence of pSMART CMV-HIF

AGATCTTGAATAATAAAATGTGTGTTTGTCCGAAATACGCGTTTTGAGATTTCTGTCGCCGACTAA
ATTCATGTCGCGCGATAGTGGTGTTTATCGCCGATAGAGATGGCGATATTGGAAAAATTGATATTT
GAAAATATGGCATATTGAAAATGTCGCCGATGTGAGTTTCTGTGTAACTGATATCGCCATTTTTCC
AAAAGTGATTTTTGGGCATACGCGATATCTGGCGATAGCGCTTATATCGTTTACGGGGGATGGCGA
TAGACGACTTTGGTGACTTGGGCGATTCTGTGTGTCGCAAATATCGCAGTTTCGATATAGGTGACA
GACGATATGAGGCTATATCGCCGATAGAGGCGACATCAAGCTGGCACATGGCCAATGCATATCGAT
CTATACATTGAATCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCAATAT
TGGCTATTGGCCATTGCATACGTTGTATCCATATCGTAATATGTACATTTATATTGGCTCATGTCC
AACATTACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT
AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACC
GCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGAC
TTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA
TCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCA
GTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCAT
GGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAG
TCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATG
TCGTAACAACTGCGATCGCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTC
TATATAAGCAGAGCTCGTTTAGTGAACCGGGCACTCAGATTCTGCGGTCTGAGTCCCTTCTCTGCT
GGGCTGAAAAGGCCTTTGTAATAAATATAATTCTCTACTCAGTCCCTGTCTCTAGTTTGTCTGTTC
GAGATCCTACAGTTGGCGCCCGAACAGGGACCTGAGAGGGGCGCAGACCCTACCTGTTGAACCTGG
CTGATCGTAGGATCCCCGGGACAGCAGAGGAGAACTTACAGAAGTCTTCTGGAGGTGTTCCTGGCC
AGAACACAGGAGGACAGGTAAGATTGGGAGACCCTTTGACATTGGAGCAAGGCGCTCAAGAAGTTA
GAGAAGGTGACGGTACAAGGGTCTCAGAAATTAACTACTGGTAACTGTAATTGGGCGCTAAGTCTA
GTAGACTTATTTCATGATACCAACTTTGTAAAAGAAAAGGACTGGCAGCTGAGGGATGTCATTCCA
TTGCTGGAAGATGTAACTCAGACGCTGTCAGGACAAGAAAGAGAGGCCTTTGAAAGAACATGGTGG
GCAATTTCTGCTGTAAAGATGGGCCTCCAGATTAATAATGTAGTAGATGGAAAGGCATCATTCCAG
CTCCTAAGAGCGAAATATGAAAAGAAGACTGCTAATAAAAAGCAGTCTGAGCCCTCTGAAGAATAT
CTCTAGAGTCGACGCTCTCATTACTTGTAACAAAGGGAGGGAAAGTATGGGAGGACAGACACCATG
GGAAGTATTTATCACTAATCAAGCACAAGTAATACATGAGAAACTTTTACTACAGCAAGCACAATC
CTCCAAAAAATTTTGTTTTTACAAAATCCCTGGTGAACATGGTCGACTCTAGAACTAGTGGATCCC
CCGGGCTGCAGGAGTGGGGAGGCACGATGGCCGCTTTGGTCGAGGCGGATCCGGCCATTAGCCATA
TTATTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTGCATACGTTGTATCCATAT
CATAATATGTACATTTATATTGGCTCATGTCCAACATTACCGCCATGTTGACATTGATTATTGACT
AGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACA
TAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATG
ACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGG
TAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAAT
GACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAG
TACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGT
GGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTT
TGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGC
GGTAGGCATGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGG
AGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGCGGCCGG
GAACGGTGCATTGGAAGCTTGGTACCGGCTAGCCGTAGAATCCGACCGATTCACCATGGAGGGCGC
CGGCGGCGCGAACGACAAGAAAAAGATAAGTTCTGAACGTCGAAAAGAAAAGTCTCGAGATGCAGC
CAGATCTCGGCGAAGTAAAGAATCTGAAGTTTTTTATGAGCTTGCTCATCAGTTGCCACTTCCACA
TAATGTGAGTTCGCATCTTGATAAGGCCTCTGTGATGAGGCTTACCATCAGCTATTTGCGTGTGAG
GAAACTTCTGGATGCTGGTGATTTGGATATTGAAGATGACATGAAAGCACAGATGAATTGCTTTTA
TTTGAAAGCCTTGGATGGTTTTGTTATGGTTCTCACAGATGATGGTGACATGATTTACATTTCTGA
TAATGTGAACAAATACATGGGATTAACTCAGTTTGAACTAACTGGACACAGTGTGTTTGATTTTAC
TCATCCATGTGACCATGAGGAAATGAGAGAAATGCTTACACACAGAAATGGCCTTGTGAAAAAGGG

TAAAGAACAAAACACACAGCGAAGCTTTTTTCTCAGAATGAAGTGTACCCTAACTAGCCGAGGAAG
AACTATGAACATAAAGTCTGCAACATGGAAGGTATTGCACTGCACAGGCCACATTCACGTATATGA
TACCAACAGTAACCAACCTCAGTGTGGGTATAAGAAACCACCTATGACCTGCTTGGTGCTGATTTG
TGAACCCATTCCTCACCCATCAAATATTGAAATTCCTTTAGATAGCAAGACTTTCCTCAGTCGACA
CAGCCTGGATATGAAATTTTCTTATTGTGATGAAAGAATTACCGAATTGATGGGATATGAGCCAGA
AGAACTTTTAGGCCGCTCAATTTATGAATATTATCATGCTTTGGACTCTGATCATCTGACCAAAAC
TCATCATGATATGTTTACTAAAGGACAAGTCACCACAGGACAGTACAGGATGCTTGCCAAAAGAGG
TGGATATGTCTGGGTTGAAACTCAAGCAACTGTCATATATAACACCAAGAATTCTCAACCACAGTG
CATTGTATGTGTGAATTACGTTGTGAGTGGTATTATTCAGCACGACTTGATTTTCTCCCTTCAACA
AACAGAATGTGTCCTTAAACCGGTTGAATCTTCAGATATGAAAATGACTCAGCTATTCACCAAAGT
TGAATCAGAAGATACAAGTAGCCTCTTTGACAAACTTAAGAAGGAACCTGATGCTTTAACTTTGCT
GGCCCCAGCCGCTGGAGACACAATCATATCTTTAGATTTTGGCAGCAACGACACAGAAACTGATGA
CCAGCAACTTGAGGAAGTACCATTATATAATGATGTAATGCTCCCCTCACCCAACGAAAAATTACA
GAATATAAATTTGGCAATGTCTCCATTACCCACCGCTGAAACGCCAAAGCCACTTCGAAGTAGTGC
TGACCCTGCACTCAATCAAGAAGTTGCATTAAAATTAGAACCAAATCCAGAGTCACTGGAACTTTC
TTTTACCATGCCCCAGATTCAGGATCAGACACCTAGTCCTTCCGATGGAAGCACTAGACAAAGTTC
ACCTGAGCCTAATAGTCCCAGTGAATATTGTTTTTATGTGGATAGTGATATGGTCAATGAATTCAA
GTTGGAATTGGTAGAAAAACTTTTTGCTGAAGACACAGAAGCAAAGAACCCATTTTCTACTCAGGA
CACAGATTTAGACTTGGAGATGTTAGCTCCCTATATCCCAATGGATGATGACTTCCAGTTACGTTC
CTTCGATCAGTTGTCACCATTAGAAAGCAGTTCCGCAAGCCCTGAAAGCGCAAGTCCTCAAAGCAC
AGTTACAGTATTCCAGCAGACTCAAATACAAGAACCTACTGCTAATGCCACCACTACCACTGCCAC
CACTGATGAATTAAAAACAGTGACAAAAGACCGTATGGAAGACATTAAAATATTGATTGCATCTCC
ATCTCCTACCCACATACATAAAGAAACTACTAGTGCCACATCATCACCATATAGAGATACTCAAAG
TCGGACAGCCTCACCAAACAGAGCAGGAAAAGGAGTCATAGAACAGACAGAAAAATCTCATCCAAG
AAGCCCTAACGTGTTATCTGTCGCTTTGAGTCAAAGAACTACAGTTCCTGAGGAAGAACTAAATCC
AAAGATACTAGCTTTGCAGAATGCTCAGAGAAAGCGAAAAATGGAACATGATGGTTCACTTTTTCA
AGCAGTAGGAATTGGAACATTATTACAGCAGCCAGACGATCATGCAGCTACTACATCACTTTCTTG
GAAACGTGTAAAAGGATGCAAATCTAGTGAACAGAATGGAATGGAGCAAAAGACAATTATTTTAAT
ACCCTCTGATTTAGCATGTAGACTGCTGGGGCAATCAATGGATGAAAGTGGATTACCACAGCTGAC
CAGTTATGATTGTGAAGTTAATGCTCCTATACAAGGCAGCAGAAACCTACTGCAGGGTGAAGAATT
ACTCAGAGCTTTGGATCAAGTTAACTGAGCGGATCCGACGGGGATCCTCTAGCGTTATCCATCACA
CTGGCGGCCGCGACTCTAGAGTCGACCTCGAGGGGGGGCCCGGACCTACTAGGGTGCTGTGGAAGG
GTGATGGTGCAGTAGTAGTTAATGATGAAGGAAAGGGAATAATTGCTGTACCATTAACCAGGACTA
AGTTACTAATAAAACCAAATTGAGTATTGTTGCAGGAAGCAAGACCCAACTACCATTGTCAGCTGT
GTTTCCTGACCTCAATATTTGTTATAAGGTTTGATATGAATCCCAGGGGGAATCTCAACCCCTATT
ACCCAACAGTCAGAAAAATCTAAGTGTGAGGAGAACACAATGTTTCAACCTTATTGTTATAATAAT
GACAGTAAGAACAGCATGGCAGAATCGAAGGAAGCAAGAGACCAAGAATGAACCTGAAAGAAGAAT
CTAAAGAAGAAAAAAGAAGAAATGACTGGTGGAAAATAGGTATGTTTCTGTTATGCTTAGCAGGAA
CTACTGGAGGAATACTTTGGTGGTATGAAGGACTCCCACAGCAACATTATATAGGGTTGGTGGCGA
TAGGGGGAAGATTAAACGGATCTGGCCAATCAAATGCTATAGAATGCTGGGGTTCCTTCCCGGGGT
GTAGACCATTTCAAAATTACTTCAGTTATGAGACCAATAGAAGCATGCATATGGATAATAATACTG
CTACATTATTAGAAGCTTTAACCAATATAACTGCTCTATAAATAACAAAACAGAATTAGAAACATG
GAAGTTAGTAAAGACTTCTGGCATAACTCCTTTACCTATTTCTTCTGAAGCTAACACTGGACTAAT
TAGACATAAGAGAGATTTTGGTATAAGTGCAATAGTGGCAGCTATTGTAGCCGCTACTGCTATTGC
TGCTAGCGCTACTATGTCTTATGTTGCTCTAACTGAGGTTAACAAAATAATGGAAGTACAAAATCA
TACTTTTGAGGTAGAAAATAGTACTCTAAATGGTATGGATTTAATAGAACGACAAATAAAGATATT
ATATGCTATGATTCTTCAAACACATGCAGATGTTCAACTGTTAAAGGAAAGACAACAGGTAGAGGA
GACATTTAATTTAATTGGATGTATAGAAAGAACACATGTATTTTGTCATACTGGTCATCCCTGGAA
TATGTCATGGGGACATTTAAATGAGTCAACACAATGGGATGACTGGGTAAGCAAAATGGAAGATTT
AAATCAAGAGATACTAACTACACTTCATGGAGCCAGGAACAATTTGGCACAATCCATGATAACATT
CAATACACCAGATAGTATAGCTCAATTTGGAAAAGACCTTTGGAGTCATATTGGAAATTGGATTCC
TGGATTGGGAGCTTCCATTATAAAATATATAGTGATGTTTTTGCTTATTTATTTGTTACTAACCTC
TTCGCCTAAGATCCTCAGGGCCCTCTGGAAGGTGACCAGTGGTGCAGGGTCCTCCGGCAGTCGTTA
CCTGAAGAAAAAATTCCATCACAAACATGCATCGCGAGAAGACACCTGGGACCAGGCCCAACACAA
CATACACCTAGCAGGCGTGACCGGTGGATCAGGGGACAAATACTACAAGCAGAAGTACTCCAGGAA

CGACTGGAATGGAGAATCAGAGGAGTACAACAGGCGGCCAAAGAGCTGGGTGAAGTCAATCGAGGC
ATTTGGAGAGAGCTATATTTCCGAGAAGACCAAAGGGGAGATTTCTCAGCCTGGGGCGGCTATCAA
CGAGCACAAGAACGGCTCTGGGGGGAACAATCCTCACCAAGGGTCCTTAGACCTGGAGATTCGAAG
CGAAGGAGGAAACATTTATGACTGTTGCATTAAAGCCCAAGAAGGAACTCTCGCTATCCCTTGCTG
TGGATTTCCCTTATGGCTATTTTGGGGACTAGTAATTATAGTAGGACGCATAGCAGGCTATGGATT
ACGTGGACTCGCTGTTATAATAAGGATTTGTATTAGAGGCTTAAATTTGATATTTGAAATAATCAG
AAAAATGCTTGATTATATTGGAAGAGCTTTAAATCCTGGCACATCTCATGTATCAATGCCTCAGTA
TGTTTAGAAAAACAAGGGGGGAACTGTGGGGTTTTTATGAGGGGTTTTATAAATGATTATAAGAGT
AAAAAGAAAGTTGCTGATGCTCTCATAACCTTGTATAACCCAAAGGACTAGCTCATGTTGCTAGGC
AACTAAACCGCAATAACCGCATTTGTGACGCGAGTTCCCCATTGGTGACGCGTTAACTTCCTGTTT
TTACAGTATATAAGTGCTTGTATTCTGACAATTGGGCACTCAGATTCTGCGGTCTGAGTCCCTTCT
CTGCTGGGCTGAAAAGGCCTTTGTAATAAATATAATTCTCTACTCAGTCCCTGTCTCTAGTTTGTC
TGTTCGAGATCCTACAGAGCTCATGCCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAAT
TGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCC
TAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTG
TCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCT
TCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCAC
TCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAA
GGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCC
CCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGA
TACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA
TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTC
AGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGC
TGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCA
GCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGG
TGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACC
TTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTT
GTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACG
GGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGG
ATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAA
ACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGT
TCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGC
CCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAG
CCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAAT
TGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCT
ACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCA
AGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTT
GTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACT
GTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAG
TGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGA
ACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTG
TTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACC
AGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGG
AAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTC
ATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCC
CGAAAAGTGCCACCTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAA
TCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCG
AGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACG
TCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTT
TTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTT
GACGGGGAAAGCCAACCTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCGGC

SEQ ID NO:43

The nucleotide sequence of pSMART CMV-empty

AGATCTTGAATAATAAAATGTGTGTTTGTCCGAAATACGCGTTTTGAGATTTCTGTCGCCGACTAA
ATTCATGTCGCGCGATAGTGGTGTTTATCGCCGATAGAGATGGCGATATTGGAAAAATTGATATTT
GAAAATATGGCATATTGAAAATGTCGCCGATGTGAGTTTCTGTGTAACTGATATCGCCATTTTTCC
AAAAGTGATTTTTGGGCATACGCGATATCTGGCGATAGCGCTTATATCGTTTACGGGGGATGGCGA
TAGACGACTTTGGTGACTTGGGCGATTCTGTGTGTCGCAAATATCGCAGTTTCGATATAGGTGACA
GACGATATGAGGCTATATCGCCGATAGAGGCGACATCAAGCTGGCACATGGCCAATGCATATCGAT
CTATACATTGAATCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCAATAT
TGGCTATTGGCCATTGCATACGTTGTATCCATATCGTAATATGTACATTTATATTGGCTCATGTCC
AACATTACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT
AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACC
GCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGAC
TTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTA
TCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCA
GTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCAT
GGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAG
TCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATG
TCGTAACAACTGCGATCGCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTC
TATATAAGCAGAGCTCGTTTAGTGAACCGGGCACTCAGATTCTGCGGTCTGAGTCCCTTCTCTGCT
GGGCTGAAAAGGCCTTTGTAATAAATATAATTCTCTACTCAGTCCCTGTCTCTAGTTTGTCTGTTC
GAGATCCTACAGTTGGCGCCCGAACAGGGACCTGAGAGGGGCGCAGACCCTACCTGTTGAACCTGG
CTGATCGTAGGATCCCCGGGACAGCAGAGGAGAACTTACAGAAGTCTTCTGGAGGTGTTCCTGGCC
AGAACACAGGAGGACAGGTAAGATTGGGAGACCCTTTGACATTGGAGCAAGGCGCTCAAGAAGTTA
GAGAAGGTGACGGTACAAGGGTCTCAGAAATTAACTACTGGTAACTGTAATTGGGCGCTAAGTCTA
GTAGACTTATTTCATGATACCAACTTTGTAAAAGAAAAGGACTGGCAGCTGAGGGATGTCATTCCA
TTGCTGGAAGATGTAACTCAGACGCTGTCAGGACAAGAAAGAGAGGCCTTTGAAAGAACATGGTGG
GCAATTTCTGCTGTAAAGATGGGCCTCCAGATTAATAATGTAGTAGATGGAAAGGCATCATTCCAG
CTCCTAAGAGCGAAATATGAAAAGAAGACTGCTAATAAAAAGCAGTCTGAGCCCTCTGAAGAATAT
CTCTAGAGTCGACGCTCTCATTACTTGTAACAAAGGGAGGGAAAGTATGGGAGGACAGACACCATG
GGAAGTATTTATCACTAATCAAGCACAAGTAATACATGAGAAACTTTTACTACAGCAAGCACAATC
CTCCAAAAAATTTTGTTTTTACAAAATCCCTGGTGAACATGGTCGACTCTAGAACTAGTGGATCCC
CCGGGCTGCAGGAGTGGGGAGGCACGATGGCCGCTTTGGTCGAGGCGGATCCGGCCATTAGCCATA
TTATTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTGCATACGTTGTATCCATAT
CATAATATGTACATTTATATTGGCTCATGTCCAACATTACCGCCATGTTGACATTGATTATTGACT
AGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACA
TAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATG
ACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGG
TAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAAT
GACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAG
TACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGT
GGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTT
TGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGC
GGTAGGCATGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGG
CCGCGACTCTAGAGTCGACCTCGAGGGGGGGCCCGGACCTACTAGGGTGCTGTGGAAGGGTGATGG
TGCAGTAGTAGTTAATGATGAAGGAAAGGGAATAATTGCTGTACCATTAACCAGGACTAAGTTACT
AATAAAACCAAATTGAGTATTGTTGCAGGAAGCAAGACCCAACTACCATTGTCAGCTGTGTTTCCT
GACCTCAATATTTGTTATAAGGTTTGATATGAATCCCAGGGGGAATCTCAACCCCTATTACCCAAC
AGTCAGAAAAATCTAAGTGTGAGGAGAACACAATGTTTCAACCTTATTGTTATAATAATGACAGTA
AGAACAGCATGGCAGAATCGAAGGAAGCAAGAGACCAAGAATGAACCTGAAAGAAGAATCTAAAGA
AGAAAAAAGAAGAAATGACTGGTGGAAAATAGGTATGTTTCTGTTATGCTTAGCAGGAACTACTGG
AGGAATACTTTGGTGGTATGAAGGACTCCCACAGCAACATTATATAGGGTTGGTGGCGATAGGGGG
AAGATTAAACGGATCTGGCCAATCAAATGCTATAGAATGCTGGGGTTCCTTCCCGGGGGTGTAGACC

ATTTCAAAATTACTTCAGTTATGAGACCAATAGAAGCATGCATATGGATAATAATACTGCTACATT
ATTAGAAGCTTTAACCAATATAACTGCTCTATAAATAACAAAACAGAATTAGAAACATGGAAGTTA
GTAAAGACTTCTGGCATAACTCCTTTACCTATTTCTTCTGAAGCTAACACTGGACTAATTAGACAT
AAGAGAGATTTTGGTATAAGTGCAATAGTGGCAGCTATTGTAGCCGCTACTGCTATTGCTGCTAGC
GCTACTATGTCTTATGTTGCTCTAACTGAGGTTAACAAAATAATGGAAGTACAAAATCATACTTTT
GAGGTAGAAAATAGTACTCTAAATGGTATGGATTTAATAGAACGACAAATAAAGATATTATATGCT
ATGATTCTTCAAACACATGCAGATGTTCAACTGTTAAAGGAAAGACAACAGGTAGAGGAGACATTT
AATTTAATTGGATGTATAGAAAGAACACATGTATTTTGTCATACTGGTCATCCCTGGAATATGTCA
TGGGGACATTTAAATGAGTCAACACAATGGGATGACTGGGTAAGCAAAATGGAAGATTTAAATCAA
GAGATACTAACTACACTTCATGGAGCCAGGAACAATTTGGCACAATCCATGATAACATTCAATACA
CCAGATAGTATAGCTCAATTTGGAAAAGACCTTTGGAGTCATATTGGAAATTGGATTCCTGGATTG
GGAGCTTCCATTATAAAATATATAGTGATGTTTTTGCTTATTTATTTGTTACTAACCTCTTCGCCT
AAGATCCTCAGGGCCCTCTGGAAGGTGACCAGTGGTGCAGGGTCCTCCGGCAGTCGTTACCTGAAG
AAAAAATTCCATCACAAACATGCATCGCGAGAAGACACCTGGGACCAGGCCCAACACAACATACAC
CTAGCAGGCGTGACCGGTGGATCAGGGGACAAATACTACAAGCAGAAGTACTCCAGGAACGACTGG
AATGGAGAATCAGAGGAGTACAACAGGCGGCCAAAGAGCTGGGTGAAGTCAATCGAGGCATTTGGA
GAGAGCTATATTTCCGAGAAGACCAAAGGGGAGATTTCTCAGCCTGGGGCGGCTATCAACGAGCAC
AAGAACGGCTCTGGGGGGAACAATCCTCACCAAGGGTCCTTAGACCTGGAGATTCGAAGCGAAGGA
GGAAACATTTATGACTGTTGCATTAAAGCCCAAGAAGGAACTCTCGCTATCCCTTGCTGTGGATTT
CCCTTATGGCTATTTTGGGGACTAGTAATTATAGTAGGACGCATAGCAGGCTATGGATTACGTGGA
CTCGCTGTTATAATAAGGATTTGTATTAGAGGCTTAAATTTGATATTTGAAATAATCAGAAAAATG
CTTGATTATATTGGAAGAGCTTTAAATCCTGGCACATCTCATGTATCAATGCCTCAGTATGTTTAG
AAAAACAAGGGGGGAACTGTGGGGTTTTTATGAGGGGTTTTATAAATGATTATAAGAGTAAAAAGA
AAGTTGCTGATGCTCTCATAACCTTGTATAACCCAAAGGACTAGCTCATGTTGCTAGGCAACTAAA
CCGCAATAACCGCATTTGTGACGCGAGTTCCCCATTGGTGACGCGTTAACTTCCTGTTTTTACAGT
ATATAAGTGCTTGTATTCTGACAATTGGGCACTCAGATTCTGCGGTCTGAGTCCCTTCTCTGCTGG
GCTGAAAAGGCCTTTGTAATAAATATAATTCTCTACTCAGTCCCTGTCTCTAGTTTGTCTGTTCGA
GATCCTACAGAGCTCATGCCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATC
CGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAG
TGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC
AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTT
CCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGG
CGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGC
AAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACG
AGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGG
CGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGT
CCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGG
TGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCT
TATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA
CTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTA
ACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAA
AAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCA
AGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTG
ACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCA
CCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGT
CTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCA
TAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTG
CTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCG
GAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCC
GGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCA
TCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAG
TTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAA
GTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGC
CATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGC
GGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAA

AAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGAT
CCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTT
CTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTT
GAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCG
GATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG
TGCCACCTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTC
ATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGG
GTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGG
GCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGG
GTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGG
AAAGCCAACCTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCGGC